International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Impression Generation from X-ray Images Using Machine Learning

Suyog Sawant[1], Rushikesh Sawant[2], Sarvesh Karpe[3], Vaishnav Kubade[4], Aditya Kedari[5]

[1,2,3,4,5]Computer Science and Engineering (AIML), SSPM College of Engineering (Mumbai University), Kankavali, India

Abstract

Due to the incorporation of machine learning techniques, medical image analysis has made significant strides in recent years. In this paper, we concentrate on a crucial application: employing machine learning to extract literary impressions from chest X-ray pictures. The goal of the project is to fill the gap between natural language processing and medical imaging by enabling the automatic creation of radio-logical impressions, which are crucial for diagnostic reports. The problem statement includes a number of significant difficulties. A lack of positive ex- amples compared to negative cases in medical datasets can skew model training and impair diagnostic pre- cision. For impression formation, accurate feature extraction from chest X-ray pictures is essential. The CheXNet model must be carefully adjusted to the particular task at hand in order to be used for this purpose. Converting visual information into cohesive literary perceptions is the main challenge. To do this while minimizing the risk of overfitting, a sequence-to- sequence model with an attention mechanism must be used to precisely match image attributes with textual information. These difficulties highlight the project's complexity and highlight the requirement for exact machine learning methods, advanced architectural layouts, and strategic feature extraction in order to enable the automated generation of high-quality radio-logical images.

Keywords: Machine learning, X-ray, Datasets, Models, Accuracy, Precision, Recall, Activation Functions.

1. INTRODUCTION

Recent years have seen notable advances in a number of fields, most notably medical image anal- ysis, because to the convergence of deep learning, natural language processing (NLP), and computer vision (CV). The work of creating literary descrip- tions, or "impressions," from medical photos is one such use that is gaining more and more attention. This procedure is commonly known as image cap- tioning. The combination of CV and NLP has the potential to completely transform medical diagnosis and treatment planning, especially in radiology. A staple of medical imaging, X-rays offer price- less insights into the internal workings of the human body and help identify and diagnose a wide range of conditions, from cancers and lung disorders to fractures and bone injuries. In the past, experienced radiologists and pathologists have been largely re- sponsible for interpreting X-ray reports. They care- fully examine both objective and subjective data in order to make precise diagnosis. Inter-observer variability and the time-consuming nature of man- ual analysis are two difficulties with this technique,

though.

In response, this study aims to improve and automate the examination of chest X-ray pictures by utilizing deep learning techniques. In particular, we tackle the picture captioning task: we train a deep learning model to produce impressions, which are descriptive textual summaries, based on input X-ray images. Our goal is to extract relevant charac- teristics from chest X-ray images and convert them into clinically meaningful textual descriptions by fusing cutting-edge deep learning architectures with transfer learning techniques.

The urgent need to improve and streamline ra- diology's diagnostic procedure in order to pro- vide patients with prompt and correct care is the driving force behind this research endeavor. We envision a time when radiologists will be able to use sophisticated computational tools to speed up workflow, lower diagnostic mistake rates, and ultimately enhance patient outcomes by automating the creation of X-ray impressions. Additionally, the suggested method shows potential for generalizabil- ity and scalability across a range of medical imaging modalities, opening the door to more widespread healthcare applications.

We outline the issue statement, experimental design, methodology, and findings of our study on picture captioning for chest X-ray diagnosis in this paper. Furthermore, we give a thorough assessment of the authenticity of generated impres- sions in comparison to human-expert annotations using recognized performance criteria, such as the bilingual evaluation understudy (BLEU) score, in our suggested deep learning model. Our goal is to further the integration of AI-driven technologies into clinical practice and make a positive impact on the rapidly developing field of medical image analysis.

2. Background And Related Work

The fields of computer vision and natural lan- guage processing have undergone a revolution be- cause to deep learning techniques, which have made significant progress possible in tasks like picture captioning. The convergence of these fields has sparked the creation of advanced models that can produce written summaries that are descriptive based on input images, allowing for better compre- hension and interpretation of visual information.

Jay Alammar's work on the attention process is a significant contribution to the field of image captioning. By simulating human visual attention, the attention mechanism enables models to dynam- ically focus on different areas of an image when generating captions, greatly enhancing the caliber and relevancy of generated descriptions.

The investigation of beam search algorithms in picture captioning, which attempts to produce sev- eral candidate captions and choose the most likely one based on a scoring system, is another notewor- thy addition. This method, which gained popularity in jobs involving natural language processing, has been modified and improved for use in image captioning applications, producing outputs that are more varied and pertinent to the context.

The practical application of deep learning tech- niques in this field is demonstrated by Harshall Lamba's work on image captioning using Keras. Using the Flickr8k dataset, Lamba combined recur- rent neural networks (RNNs) with attention mech- anisms for caption generation with pre-trained con- volutional neural networks (like Inception v3) for feature extraction from photos.

Furthermore, Lamba emphasizes the significance of carefully selecting and fine-tuning model pa- rameters to obtain optimal performance through his selection of optimization strategies, which include the use of pretrained GloVe word embeddings, the Adam optimizer, and the categorical cross-entropy

loss function. The iterative nature of model op- timization and refining in deep learning tasks is highlighted by his experiments with various hyper- parameters, including batch size and learning rate. To sum up, the research contributions stated above offer significant perspectives and approaches for addressing the job of picture captioning. This establishes the groundwork for our own study on utilizing deep learning to produce impressions from chest X-ray images. Building on previous efforts, our ultimate goal is to improve patient care and diagnostic accuracy in clinical practice by investi- gating new strategies and optimizations suited to the unique demands and difficulties of medical image processing.

3. METHODOLOGIES

In this section, we mentioned various methodologies that we have learned and done the analysis of ML Algorithms such as SVM, Random Forest, Naive Bayes, CNN, XGBoost, Decision Tree, and Feature Selection. The proposed framework in this paper, shown in the figure, used machine learning algorithms with selected features and datasets. The model mainly used three phases, named A) data acquisition, B) data preprocessing, and C) website classification for detecting phishing websites. Acquired data were preprocessed and fed into the feed-forward neural network. A performance evaluation matrix (confusion matrix) was used to evaluate the performance of the machine learning approach to visualize the experiments.

A. Algorithms

1. **Chexnet::** CheXNet is a convolutional neu- ral network (CNN) model that utilizes the DenseNet121 architecture and is tailored for med- ical image processing. From the chest X-ray pic- tures, useful features are extracted using CheXNet, identifying pertinent patterns and structures neces- sary for precise analysis. The top layer, which is usually employed for classification tasks, is ignored in favor of concentrating only on feature extraction once the model's weights have been set. Every chest X-ray image is preprocessed to conform to DenseNet121's input specifications. The retrieved features, which stand for significant attributes in the pictures, are then used as input for the following parts of the picture captioning model, providing essential data for producing reports with descriptive text.
2. **Greedy search::** Based on the model's predic- tions, this straightforward and effective approach is used to produce captions in sequential order. Dur- ing the caption generating process, Greedy Search chooses the word from the lexicon that has the highest likelihood at each time step. The follow- ing output token, which goes toward creating the final caption, is this chosen word. Greedy Search helps create captions that are both coherent and contextually appropriate by iteratively selecting the most likely terms at each time step. This method guarantees that the resulting captions accurately describe the radiological findings seen in the chest X-ray images, effectively conveying the information and context captured by the image characteristics.
3. **LSTM::** In the project, the decoder stage of the picture captioning model heavily relies on the Long Short-Term Memory (LSTM) architec- ture. Recurrent neural networks (RNNs) of the long-term dependency (LSTM) type are specifically engineered to capture long-term dependencies in sequential input. The encoded image features that were taken out of the chest X-ray pictures are processed by LSTM using its special architecture. The LSTM network receives these features, which stand for the significant traits and patterns seen in the photos. At the decoder stage, LSTM uses the contextual data from the encoded image character- istics to produce word-by-word output

captions in a sequential fashion. The coherence and contextuality of the generated textual descriptions are guaranteed by LSTM, which efficiently captures the linkages and dependencies between succeeding words in the output captions.

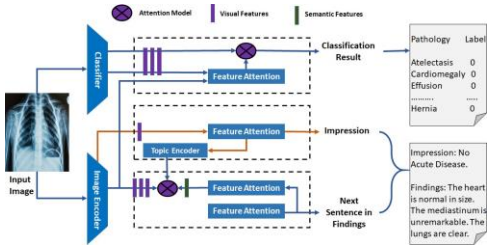

Fig. 1. Proposed System

B. Proposed System

1. Data Collection and Preprocessing:

Acqui- sition of Image Data: Download frontal and lateral views of chest X-ray pictures in the.png format from Indiana University's dataset. Radiology Report Retrieval: Access related radi- ological reports that are saved in XML format and include written summaries of clinical observations. **Preprocessing:** Remove extraneous tags, punctu- ation, and numbers from text data by cleaning and preprocessing it. Standardize contractions and ad- dress missing values by removing cells with blank picture names and inserting placeholders for NaN values in text data. Aligning image-text pairs and resolving any inconsistencies will help to ensure data consistency and integrity.

2. Exploratory Data Analysis (EDA):

Image Data Exploration: To learn more about the composi- tion of datasets, examine image properties including size, resolution, and content. Text Data Analysis: The 'Findings' feature should be the main focus of the case study's target variable. Evaluating the word count distribution in 'Findings' entries will help you comprehend the length and intricacy of radiology reports. To uncover important medical terminology and abnor- malities emphasized in radiological findings, use wordcloud visualization.

3. Data Structure and Feature Engineering:

Image Feature Extraction: To extract high-level fea- tures from chest X-ray pictures, use convolutional neural networks (CNNs) that have already been trained, like Inception v3.Text Data **Processing:** To maintain consistency in input data, tokenize radiological reports and use strategies like padding. To represent textual data in a continuous vector space, use word embeddings, such as GloVe vectors.

Handling Structured Data: Provide a solid strat- egy to deal with different picture counts for each patient while maintaining data structure uniformity and consistency.

4. Model Development:

Hybrid Architecture: Create a deep learning model architecture that blends recurrent neural networks (RNNs) with at- tention mechanisms for text production and CNNs for the extraction of picture features.

Method of Training: Utilizing optimization meth- ods like the Adam optimizer and the categorical cross-entropy loss function, train the model on the prepared dataset.

Hyperparameter tuning: To maxi- mize model performance, play around with various hyperparameters, such as batch size and learning rate. Evaluation Metrics: Evaluate the fidelity of gen-

erated impressions in comparison to human-expert annotations by using performance metrics, such as the bilingual evaluation understudy (BLEU) score.

5. Model evaluation and validation:

Cross-Validation: To verify model performance and guar- antee resilience, use cross-validation approaches.

Evaluation: Apply defined metrics to a different validation dataset to gauge the trained model's accuracy in predicting radiological results from chest X-ray pictures.

Comparison: Evaluate how well the suggested model performs in image captioning and medical image analysis in comparison to baseline techniques and current state-of-the-art methodologies.

C. Methods

1. Feature Extraction: Feature extraction refers to the process of transforming raw data into numer- ical features that can be processed while preserving the information in the original data set. [1] [2] There are various feature extraction techniques out of which we used PCA.

Let X be the raw data, and let $\phi(X)$ represent the extracted features. The feature extraction equation can be defined as:

$$\phi(X) = f(X_1, X_2, \ldots, X_n) \quad (1)$$

where $X_1, X_2, \ldots, X_n$ are the individual compo- nents or attributes of the raw data X, and f is the feature extraction function.

2. Feature Selection: The technique of selecting a subset of the most relevant features in the dataset to the problem is called feature selection. [3] Fea- ture selection helps ML and DL algorithms to learn more efficiently and effectively since it uses less memory and reduces time complexity, which is one of the main aims of this study. We focused on the following points for better selection.

- A feature dataset should not be a constant or should have a certain variant level.
- A feature should be correlated with the target, or it does not have any contribution to the target estimation.
- Features should not be highly correlated, or one of them does not offer any additional information than the other ones. It can only add sampling noises at this point.

$$\text{Score}(X_{\text{selected}}) = \text{Model\_Performance}(X_{\text{selected}}) \quad (2)$$

Xselected represents the selected subset of features.

Model_Performance is the performance metric of the machine learning model.

3. Activation Functions: These are the functions used to decide whether to activate or fire neurons according to input and bias. An essential part of artificial neural networks is activation functions. They give the network non-linearities, which enable it to discover intricate patterns and connections in the data. These are a few typical brain activation functions.

Tanh Activation Function:- The hyperbolic tan- gent function, commonly denoted as tanh(x), is often

used as an activation function in neural net- works. It squashes input values to the range of [1, 1]
. Here's a simple LaTeX code snippet to plot the graph of the tanh activation function.

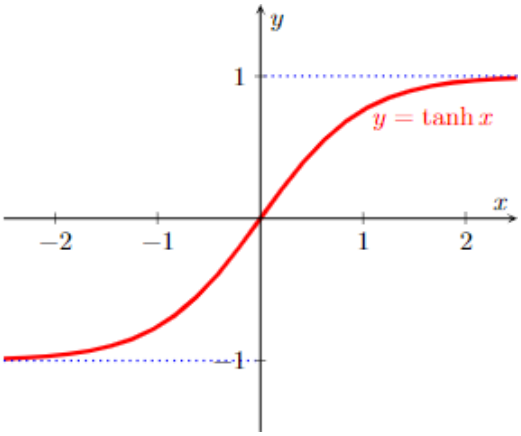

Fig. 2. Tanh Function

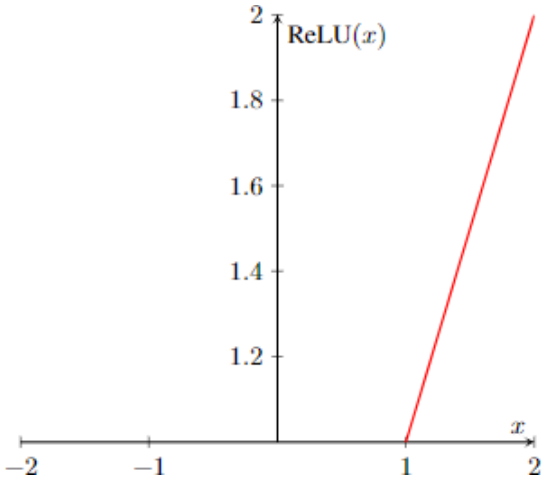

Fig. 3. ReLu Function

Relu Activation Function:-The Rectified Linear Unit (ReLU) activation function is a simple piecewise function commonly used inneural networks. It is defined as: ReLU(x)= max(0,x) In simpler terms, it outputs the input value if it is positive and zero otherwise. Here's a simple explanation of the ReLU activation function.

4. Feature Explanation: Best feature selection is an essential and challenging task for the right prediction and detection from a new dataset. They provide information about the websites that were ranked as (-1) for a phishing website, (0) for a suspicious website, and (1) for a legitimate website. Ten features of the dataset are described in Table 2 below:

TABLE I FEATURE EXPLANATION OF THE DATASET

Feature	Description
popUpWindow	Window of the web browser used for

	adjusting screen size display options for popping up the menu bar from the website.
SSLfinal State	Secure link established between client and server for communication. SSL fi- nal state is stored in the computer's cache.
Request URL	Objects like images, videos, and con- tent loaded into a web page from an- other URL.
URL of Anchor	Similar to Request URL. Phishing in- dication if different domain names are shown.
Web traffic	Defines the number of visitors to the website. Calculated in e-commerce or personal websites to attract visitors.
URL Length	Length of URL address. Less than 54 characters is considered valid, other- wise phishing or suspicious.
age of domain	Age of the domain indicates website va- lidity. Fraudulent websites have a con- cise time domain.
having IP	Valid websites generally use DNS in- stead of IP address. Phishing if an IP address is shown in the web address.
Result	All occasions sorted as "1" for "Legit- imate," "0" for "Suspicious," and "-1" for "Phishing".

4. PERFORMANCE METRICS

Various Equations are studied for classification among different ML Models. These equations such as Precision, accuracy, and Recall helped to analyze the training and testing accuracy of models. Some of them are listed below.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

Accuracy:

In machine learning, accuracy is a measure that tells us how well a model is performing. The accuracy of a model is calculated by comparing the number of correct predictions to the total number of predictions.

Precision:

Precision in machine learning is a measure that tells us how accurate a model is when it predicts a positive outcome.

Recall:

In machine learning, recall is a measure of how well a model can find and identify all the relevant instances within a dataset.

Deep Learning Based Model: Multilayer Perceptron (MLP) is the best and succesful model in the field of deep learning. It has multiple layers of perceptron with a non-linear activation function rather than a single-layer perceptron; that's why it is called a multiple-layer perceptron. It uses a backpropagation algorithm for supervised learning. A fully connected multi-layer neural network is called a Multilayer Perceptron (MLP). It has 3 layers including one hidden layer. An MLP is a typical example of a feedforward artificial neural network. If it has more than 1 hidden layer, it is called a deep ANN.

RESULTS

In this experiment performed using SVM, Ran- dom Forest, Logistic Regression, Decision Tree, and KNN using dataset 1 we got accuracy and precision as mentioned below. We got the highest accuracy and precision from the XGBoost model.

TABLE II PERFORMANCE OF ML MODELS ON "PHISHING LEGITIMATE WEBSITE" DATASET

Model	Accuracy	Precision
Logistic Regression	0.80	0.80
Decision Tree	0.81	0.81
Random Forest	0.82	0.87
KNeighborsClassifier	0.82	0.81
XGBoost	0.87	0.89
SVM	0.85	0.83

In this experiment performed using SVM, Ran- dom Forest, Logistic Regression, Decision Tree, and KNN using dataset 2 we got accuracy and precision as mentioned below. We got the highest accuracy from KNN and the highest precision from the XGBoost model.

TABLE III PERFORMANCE OF ML MODELS ON "PHISHING" DATASET

Model	Accuracy	Precision
Logistic Regression	0.80	0.80
Decision Tree	0.89	0.86
Random Forest	0.89	0.86
KNeighborsClassifier	0.92	0.82
XGBoost Classifier	0.82	0.91
SVM	0.81	0.84

In this experiment performed using SVM, Ran- dom Forest, Logistic Regression, Decision Tree, and KNN using dataset 3 we got accuracy and precision as mentioned below. We got the highest accuracy from the XGBoost model and the highest precision from the Random Forest model.

TABLE IV PERFORMANCE OF ML MODELS ON "WEBSITE PHISHING" DATASET

Model	Accuracy	Precision
Logistic Regression	0.84	0.87
Decision Tree	0.86	0.89
Random Forest	0.86	0.89
KNeighbors Classifier	0.77	0.74
XGBoost	0.90	0.86
SVM	0.88	0.86

CONCLUSION

We have examined every machine learning tech- nique that was previously covered, and the result is a model that can recognize phishing websites. Data is classified using parameters such as precision, re-call, and accuracy. Our models are sufficiently pre- dictive because they are trained on highly saturated and accurate data that is downloaded from open sources and sites like PhishTank and Kaggle.com.In order to provide the model with meaningful and trustworthy training data, we used the most modern techniques for feature extraction and selection. We were able to generate output with greater precision thanks to this method. Our objective was to use the XGBoost classifier to create a dataset with extracted URL information and transform it into 0 and 1 values and we successfully achieved the highest accuracy using the XGBoost classifier model. XG- Boost algorithm becomes very efficient in detecting phishing websites with the highest accuracy.

REFERENCES

1. Asadullah Safi and Satwinder Singh. A systematic literature review on phishing website detection techniques. Journal of King Saud University - Computer and Information Sciences, 35, 01 2023.
2. Rachael Lininger and Russell Dean Vines. Phishing: Cutting the Identity Theft Line. Wiley, 2005.
3. Ilker Kara, Murathan Ok, and Ahmet Ozaday. Character- istics of understanding urls and domain names features: The detection of phishing websites with machine learning methods. IEEE Access, 10:124420–124428, 2022.