IPTEK The Journal of Technology and Science, 33(2), 2022 (e/pISSN:2088-2033, 0853-4098)

DOI: 10.12962/j20882033.v33i2.13572 | Received 8 July, 2022; Revised 26 July, 2022; Accepted 30 July, 2022

ORIGINAL RESEARCH

EXPOSURE FUSION FRAMEWORK IN DEEP LEARNING-BASED RADIOLOGY REPORT GENERATOR

Hilya Tsaniya | Chastine Fatichah* | Nanik Suciati

Dept. of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Correspondence

*Chastine Fatichah, Dept of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. Email: chastine@if.ts.ac.id

Present Address

Gedung Teknik Informatika, Jl. Teknik Kimia, Surabaya 60111, Indonesia

Abstract

Writing a radiology report is time-consuming and requires experienced radiologists. Hence a technology that could generate an automatic report would be beneficial. The key problem in developing an automated report-generating system is providing a coherent predictive text. To accomplish this, it is important to ensure the image has good quality so that the model can learn the parts of the image in interpreting, especially in medical images that tend to be noise-prone in the acquisition process. This research uses the Exposure Fusion Framework method to enhance the quality of medical images to increase the model performance in producing coherent predictive text. The model used is an encoder-decoder with visual feature extraction using a pre-trained ChexNet, Bidirectional Encoder Representation from Transformer (BERT) embedding for text feature, and Long-short Term Memory (LSTM) as a decoder. The model's performance with EFF enhancement obtained a 7% better result than without enhancement processing using an evaluation value of Bilingual Evaluation Understudy (BLEU) with n-gram 4. It can be concluded that using the enhancement method effectively increases the model's performance.

KEYWORDS:

Exposure Fusion Framework, ChexNet, Medical Report Generator, LSTM, Learning-Based Radiology

1 | INTRODUCTION

In the last decade, the diagnostic process based on medical images is estimated to be more than 1 billion activities per year. This number continues to increase, especially in developing countries. With advancements in medical imaging technology, the diagnostic procedure based on image data is becoming more complex, resulting in an increasing need for medical professionals. Based on data from medical journals in the field of radiology in 2015, the workload of radiologists increased by 26% from the previous decade. With medical developments, radiologists need to make various comparisons to existing data to get more precise and detailed diagnoses compared to the previous years[1]. Radiologists need to compare and consider multiple factors to determine the correct diagnosis from medical images for writing good medical reports.

The increased workload would result in errors in medical reports, especially in poor conditions or fatigue. Based on professional observations from Alomedika in 2019, errors in medical reports happen with a frequency that cannot be ignored. Errors in medical reports would cause mistakes in the patient's final diagnosis. Especially if radiologists do not have sufficient experience, errors with the diagnoses are more likely to happen.

This problem with the increased workload and error frequency indicates the necessity for developing automated medical image description technology that may efficiently minimize radiologists' efforts by providing information from medical image input. An automated medical report writing system can assist radiologists in drafting reports, especially in developing countries with low resources in radiology. Radiologists can use the auto-generated descriptions to compare manual analyses, provide additional information in observing patient diagnoses, and help with diagnostic decisions.

Deep learning techniques have been used in image captioning to translate natural photos into captions similar to human interpretations. The encoder-decoder architecture underpins the deep learning approach to painting captioning, with the encoder as a technique of extracting visual features and the decoder producing text for image interpretation. Typically, the deep learning approach uses Convolutional Neural Network (CNN)[2] and Recurrent Neural Network (RNN)[3] as the encoder and decoder.

In the existing approach, several CNN variants are used as encoders in images captioning, such as Visual Geometry Group Network (VGGNet), Mobile Network[4, 5], Residual Network (ResNet), and Inception. Meanwhile, the RNN variant decoders such as Gated Recurrent Unit (GRU)[6], Long-short Term Memory (LSTM), and Bidrectional Long-short Term Memory (BLSTM)[7] become the standard decoder method in sentence prediction. To compute efficiency features, existing research also uses pre-trained modes such as VGGNet[8], ResNet[9], and Inception-V3[10].

The development of image captioning in the medical field has several challenges, including limited datasets and medical images that tend to have noise in the acquisition process. Due to this, utilizing existing captioning datasets to train the model is insufficient. The current transfer learning method with natural image data, such as ImageNet[11], still does not produce satisfying results.

In radiographic images, what tends to happen is uneven contrast. Some parts of the image are darker or lighter with uneven contrast, making it difficult for the model to learn and interpret the image. Several methods exist for adjusting the contrast to improve image quality, such as Histogram Equalization (HE), Contrast Limit Adaptive Histogram Equalization (CLAHE), and most recently, Exposure Fusion Framework (EFF)[12].

This study developed medical image captioning using an encoder-decoder-based deep learning method. To improve feature information from existing data, a pre-trained ChexNet model, which had been trained on 100k lung X-ray data[13], was used. Furthermore, an enhancement was made using EFF method to overcome the contrast problem. In addition, the BERT embedding method would enhance the words' context relevance for the text feature extraction process to boost the model's performance and generate coherent predictive text.

2 | PREVIOUS RESEARCHES

One of the first approaches to medical image captioning was made by Duygulu in 2002 by studying a collection of words associated with an image using the lexicon method[14] and then labeling radiological images using the Support Vector method. The results of the tests were also carried out through qualitative measurement based on the opinion of radiologists[15].

In a subsequent development, the encoder-decoder architecture for image captioning was first proposed in 2014 to describe sentences based on visual features of the image using the CNN-RNN method as the encoder-decoder[6, 16]. This architecture was tested in the medical field for the first time following the chest X-ray image dataset collection from Indiana University with a GoogleNet encoder and an LSTM decoder[17]. In others studies, various model of convolution encoder architectures implemented, namely Inception-v3[18], VGGNet[19], Inception-v3 with the addition of an LSTM layer on the decoder[20], ResNet-152[21], and VGG-16[22]. These researches mainly use transfer learning in the formulation of the algorithm, which still has weaknesses in the limitations of the image caption results with different preprocessing and parameters used.

To overcome the limitation, a few researchers did some preprocessing to the image, such as multiple preprocessing with augmentation[19] and object detection[21]. These are to improve image feature quality to increase model performance. Related to

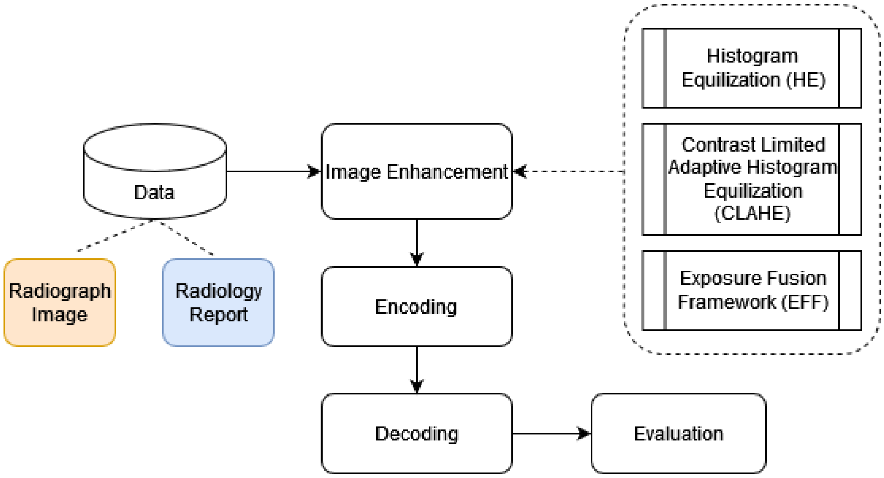


FIGURE 1 The architecture of our proposed solution.

solving uneven contrast in the image, a few methods have been invented. Histogram Equalization has proven suitable for improving contrast in an image by spreading out the pixel intensity, a variant of it, CLAHE, also give a good result with limited contrast amplification, and one of the newest method related to contrast adjustment is Exposure Fusion Framework (EFF)[12] with a weight fusion concept for better enhancement result. Applying natural images with uneven contrast shows EFF performs better than other methods[23].

3 | MATERIAL AND METHOD

The two stages of image captioning are extracting or encoding image features and generating natural language captions[24]. In this research, to improve model performance to interpret radiograph images that tend to have uneven contrast, HE, CLAHE, and EFF methods were used to enhance and compare the image quality. As an encoder to encode and extract image features, a pre-trained ChexNet was trained on similar data to the dataset. A feature vector from the encoder was used as an input to the LSTM decoder to generate sequence words.

The CNN-based model is used to extract and generate the convolutional features from the images, and each feature vector denotes a pertaining area of the image. These extracted features are then passed to the LSTM as a decoder. On the other hand, to extract text features, BERT embedding is used to increase contextual information of the text. Using LSTM as a decoder, the extracted features would be used as an input along with the feature from embedded text to generate sequence texts. The schema of the research methodology can be seen in Figure 1 .

3.1 | Dataset

This research's data is public data from Indiana University Hospital Link collection. Data contained 9199 chest X-ray images 3973 medical reports in XML format[25]. Medical reports in the dataset consist of comparison and indication, which are comparative and indication data based on the patient's symptoms and medical history, the finding is an observational analysis of the radiograph image, and the impression is the patient's diagnosis.

From the dataset exploration, one medical report can be associated with several images or not at all. Basic preprocessing such as resize and lowercase was applied to the image and text. As the reports contain more majority conditions than others, sampling was carried out on the data before splitting it into the train, validation, and test data.

3.2 | Histogram Equalization (HE)

Histogram Equalization is an enhancement method that adjusts the brightness and contrast of dark and low contrast images to improve image quality [26]. This method redistributes the gray level of the image, which can enhance the clarity of the image and result in a more equally distributed histogram. The histogram on the image can be stated as a discrete function in Eq. 1.

$$h(r_k) = n_k \tag{1}$$

Where:
h: is the symbol for histogram function,
r: is the intensity value in image pixel k,
n: is the number of the pixel with the intensity value r in the image

The histogram value was normalized by the number of pixels in the image. The final occurrence of the intensity value in the image pixel is calculated as Eq. (2).

$$p(r_k) = \frac{n_k}{I * J} \tag{2}$$

Where:
p: is the probability of occurrence after normalization
I: is the image row
J: is the image column

The final occurrence of the pixel value in the image should be scattered more evenly after normalization. This would improve the clarity of the gray image.

3.3 | Contrast Limited Adaptive Histogram Equalization (CLAHE)

CLAHE is a variant of improved HE. It applies histogram equalization over small areas of the image, enhancing the contrast of each area separately. Rather than using the image's overall information, it enhances local contrast and edges in each region according to the regional distribution of pixel intensities. Images enhanced using CLAHE appear more natural than those created using HE. A threshold parameter controls how much contrast enhancement can be performed within the chosen zone. First, the original image is changed from RGB (red, green, and blue) to HSV (hue, saturation, and value) color space to rend the color more closely related to how people see color. Second, CLAHE processes the value portion of HSV without modifying the hue or saturation. Each gray level in the first histogram is divided among the pixels in the trimmed histogram. Each pixel's value is decreased to a user-selectable limit. The HSV-processed image is then converted once more to RGB color space.

3.4 | Exposure Fusion Framework

Exposure Fusion Framework is a method to enhance image contrast using the exposure ratio in the image. This method algorithm provides better accurate contrast enhancement than other methods that usually under or over-enhanced the contrast in the image [12]. This framework addresses the issue of under- and over-enhancement with the multi-exposure in the image to obtain results with less contrast and lightness distortion. The algorithm works by fusing all pixel regions with under-exposed and over-exposed to become well-exposed pixels stated in Eq. 3.

$$R^c = \sum_{i=1}^{N} W_i \circ P_i^c \tag{3}$$

Where:
R: is the enhanced result
N: is the number of the image

P is the pixel value of the i-image
W: is the weight map of the i-image
c: is the index of color channel

Poorly exposed pixels would be given a small weight, whereas well-exposed pixels receive significant weight. The weight then is normalized, which causes the different exposure in the image region. To encounter the other exposure settings, brightness function, Brightness Transform Function (BTF) are used to differentiate exposure between images. Given an exposure ratio, the input image P is mapped as formulated in Eq. 4.

$$P_i = g(P, k_i) \tag{4}$$

Where:
P: is the pixel value of the i-image
g: is the BTF function
k: is the exposure ratio between images

The input image was combined with another exposure to simplify complexity to enhance image contrast. As the Eq. 5, the estimation of W, g, and k can be separated into three sections, which make up the enhancement problem. The whole process algorithm can be seen in Eq. 5.

$$R^c = W \circ P^c + (1 - W) \circ g(P^c, k) \tag{5}$$

The design of an enhancement algorithm that can improve the poor contrast of underexposed parts while maintaining the difference in well-exposed regions depends on the method of W. The well-exposed pixels must be given large values, and under-exposed pixels with moderately low weight values must be displayed. Highly lighted areas should be given high weight values to sustain contrast because they are highly likely to be well-exposed. The weight value for this can be calculated as Eq. 6.

$$W = T^{\mu} \tag{6}$$

Where:
T: is the map of illumination in the image
μ: parameter to control enhance the degree

To obtain an estimated illumination scene map T, optimization is required by fine-tuning T using the optimization equation in Eq. 7 to keep T's smoothness while minimizing the difference between the initial map and the refined map. The brightness component, as specified in Eq. 7, is used for the initial estimation of illumination for each pixel in the image.

$$L(x) = \max_{C \in \{R,G,B\}} P_c(x) \tag{7}$$

Eq. 7 is the brightness component symbolized with L for initial estimation illumination. To optimize the refined map, T use Eq. 8.

$$\min_T \|T - L\|_2^2 + \lambda \|M \circ \bar{V} T\|_1 \tag{8}$$

Where $\|*\|_2$ and $\|*\|_1$ are the L_2 and L_1 norm for illumination pixel in the refined value and the initial illumination value, respectively. The first order derivative filter $\bar{V}$ contains $\bar{V} h_T$ (horizontal) and $\bar{V} v_T$ (vertical) from the weight matrix. The coefficient is λ, with M is the weight matrix. The first term of this equation is to minimize the difference between the initial map and

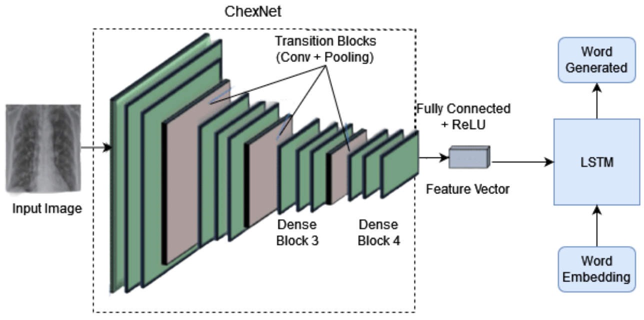


FIGURE 2 The architecture of encoder-decoder.

the refined map, while the second term maintains the smoothness of the refined map. The improvement of the lighting map's design is crucial. An enormous edge produces more similar-direction gradients in a local window than textures with intricate patterns. As a result, the weight in a window with valuable advantages should be lower than in a window that merely has textures. The formula of the weight matrix is written in Eq. 9.

$$M_d(x) = \frac{1}{|\sum_{y \in \omega(x)} \bar{V}_d L(y)| + \epsilon} \tag{9}$$

Where ϵ is a tiny constant to prevent the numerator from zero, $\omega(x)$ is the local window centered at the pixel x, and $||$ is the absolute value operator. To reduce the complexity of the method in Eq 10 (below the minimum value) and solve the function containing quadratic terms, as observed. Let x, y, z, and p stand for the corresponding vectorized forms of X, Y, Z, and P. The answer can then be achieved immediately by solving the following linear function in Eq. 11.

$$\min_T \sum_x \left((T(x) - L(x)\right)^2 + \lambda \sum_{d \in \{h,v\}} \frac{M_d(x)\, \bar{V}_d\, T(x)}{|\bar{V}_d\ L(x)| + \epsilon} \tag{10}$$

$$(I + \lambda \sum_{d \in \{h,v\}} (D_d^T\ Diag(m_d \oslash (|\bar{V}_d\ 1|) + \epsilon))\ D_t)t = 1 \tag{11}$$

Where D_d in Eq. 10 and Eq. 11 are the Toeplitz matrices or the diagonal matrix with the constant value from the discrete gradient operations with a forward difference and also the element-wise division, I is the unit matrix, the operator $Diag(v)$ creates the diagonal matrix using vector v, and the design of the weight matrix M is the primary distinction between the illumination of the map estimate approach to adjust the contrast and enhance the image.

3.5 | Image Encoder

ChexNet is a dense convolutional network (DenseNet) with 121 layers trained on 14 chest X-ray datasets. DenseNet improves information and gradients throughout the network, optimizing current networks[27]. The ChexNet model has been trained with more than 100,000 chest X-ray images on the Chest-Xray14 dataset[13]. The architecture of ChexNet as an encoder can be seen in Fig. 2 .

Chexnet is used as an encoder in this research to extract images as convolutional features from the data after enhancement. The weight from the pre-trained Chexnet is passed to the image data through the model to get the image feature vector used to generate predictive text with LSTM as the decoder.

TABLE 1 The data quantity used in modeling.

Data	Quantity
Train	3.605
Validation	901
Test	566

3.6 | Decoder

Long Short-Term Memory (LSTM) is used as a decoder. LSTM has three gates: forget gate, input gate, and output gate. The gates have their respective functions in collecting, classifying, and processing data[28]. An input modulator is used by an LSTM cell to modulate the input to the memory while it continuously learns the weights from the input gate. The weights to the output gate are also being learned through this memory cell. Each time step, the LSTM deletes, and stores some data that is subsequently used in the following time step. BERT embedding is used for the text feature to increase the contextual information in selecting text prediction.

3.7 | Evaluation

The description of the image generated by the model was evaluated using the Bilingual Evaluation Understudy (BLEU) method. The BLEU method approach counts word occurrences from the text generated by the model with the reference text from the data[29].

BLEU score using N-gram calculation is the process of calculating the match between the words generated by the model and the actual data. The N value measures precision between references sentence and predicted sentences per N given. The equation to calculate the BLEU score can be seen in Eq. 12.

$$BLEU = BP \cdot \exp \sum_{n=1}^{N} w_n \log p_n \tag{12}$$

Where:
p_n = the precision of the predictive text per n-word against the reference text,
w_n = the weight of the predictive text on the n-word,
N = gram value,
BP = penalty value of error prediction text to reference text

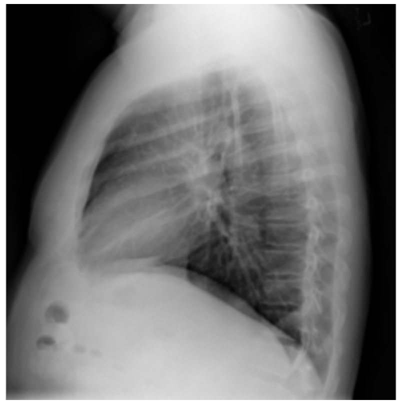
Whereas BP is a brevity punishment from the mistakes of predicted text, p is the precision value from the predicted text, w is the weight value from the words, and N is the value of the gram used to measure similarity. The value of gram used was stated as N, with a standard value of 4 to measure precision between predictive text and reference text per 4 words. In simple terms, the BLEU score is calculated as BLEU-N, with the N-gram value determining how many words are compared to the match between the predictive sentence and the reference sentence. BLEU-1 was evaluated based on matched words between the predictive sentence and the reference sentence per 1-word, BLEU-2 per 2-words, BLEU-3 per 3-words, and BLEU-4 per 4-words.

4 | RESULTS AND DISCUSSION

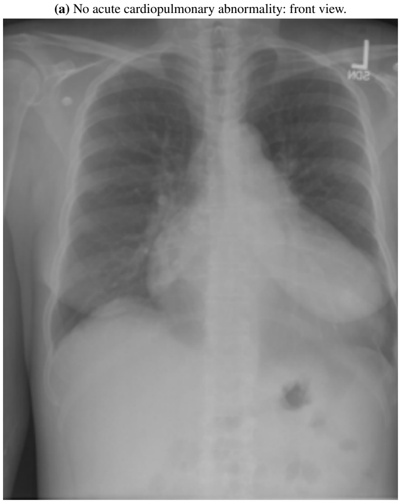
4.1 | Data Preparation

From the exploration of the extracted data, there was an imbalance between the diagnosis of x-ray images with normal conditions and several abnormal diagnoses. This can lead to overfitting the model to normal diagnostic data by writing diagnoses that tend to be the same. To overcome this, the duplicated data for a similar diagnosis is reduced, followed by data sampling to reduce the imbalances. A preprocessing is done to the medical report. The preprocessing consist of word deconstruction, deletion of characters and numbers, and conversion of letters into lowercase form. It is embedded using pre-trained BERT to get word and sentence feature extraction from text data to increase contextual information.

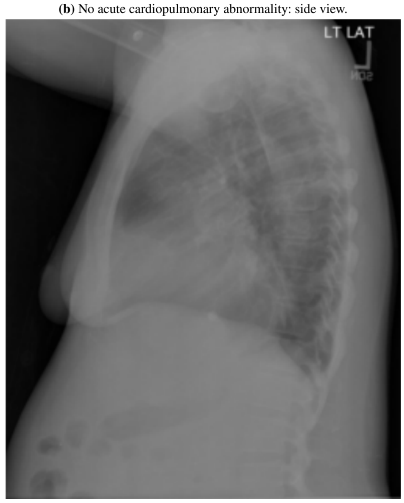
(a) No acute cardiopulmonary abnormality: front view.



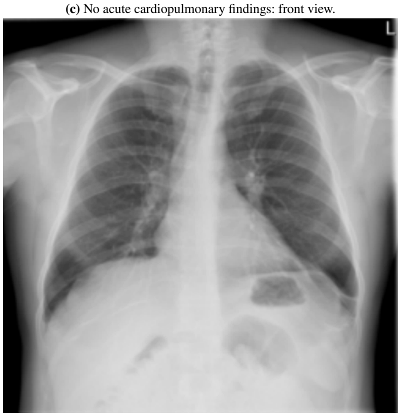
(b) No acute cardiopulmonary abnormality: side view.



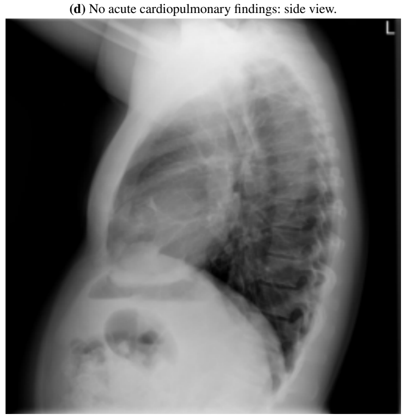
(c) No acute cardiopulmonary findings: front view.



(d) No acute cardiopulmonary findings: side view.



(e) No evidence of acute cardiopulmonary proces: front view.



(f) No evidence of acute cardiopulmonary process: side view.

FIGURE 3 Sample of data from Indiana University dataset[25]

For sampling, the data is divided by the number of report occurrences, and the data with the highest frequency is under-sampled. In contrast, the data with the lowest frequency is oversampled. The data is then separated into train, validation, and test data based on each minor and major data percentage. Table 1 shows the quantity of data used in the train, validation, and test phase.

The final data frame used for model training consists of a radiology report and the radiograph images associated with the report. For the effectivity, the radiograph in the data frame will be a pair of two images as the most frequent image number associated with a report is two. A pre-trained encoder will extract the visual feature from two images associated with the report. The sample of data that will be used for model training can be seen in Fig. 3 .

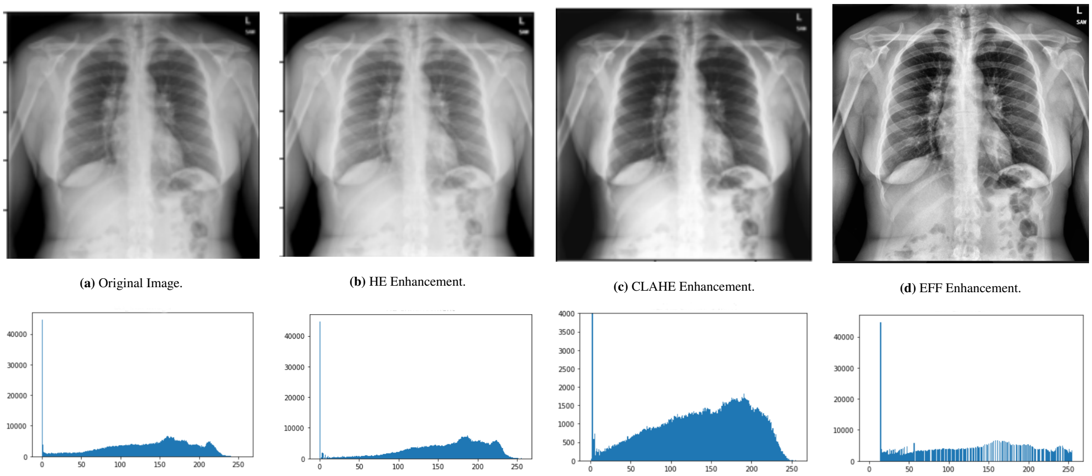


FIGURE 4 Sample of data from Indiana University dataset[25]

TABLE 2 Result comparison using deep learning method and enhancement addition.

Model	BLEU Evaluation			
	BLEU-1	BLEU-2	BLEU-3	BLEU-4
ChexNet-LSTM	0.25567	0.25576	0.290774	0.321575
ChexNet-LSTM + HE	0.22783	0.28682	0.291904	0.332374
ChexNet-LSTM + CLAHE	0.25893	0.29327	0.323290	0.337827
ChexNet-LSTM + EFF	0.25892	0.30681	0.328790	0.344820

4.2 | Enhancement Results Comparison

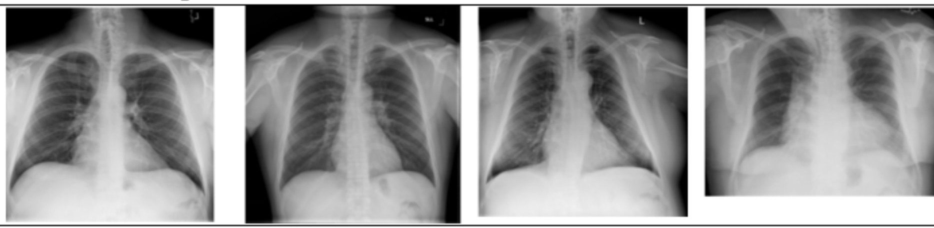
The radiograph images were enhanced using several methods, namely HE, CLAHE, and EFF. These methods are compared to see which one give better result in improving model performance. The difference in implementation of each method to the original radiograph image can be seen in Fig. 4 with the histogram graph.

From Fig. 4 , it can be seen that the implementation of the enhancement method has an impact on the radiograph image. From a visual comparison, implementing the EFF method produces better detail and contrast in the image than HE and CLAHE. Comparisons are also made based on the histogram graph of the image before and after the implementation of the enhancement method. The difference in histogram graphs shows that the enhancement method's performance changes the image's histogram graph, with the evenest distribution of pixel values obtained by the EFF method. From the comparisons, the EFF enhancement method gives the best visualization and pixel value normalization results, among other methods.

4.3 | Result

The experiments were conducted on four scenarios to compare the effect of image enhancement on the evaluation results. The first scenario, with the same dataset using an encoder-decoder architecture ChexNet and LSTM without any enhancement, then testing with the addition of enhancement HE, CLAHE, and EFF method to the image preprocessing. The results and comparison of the trials on the four scenarios can be seen in Table 2 .

Table 2 , based on the results of the BLEU evaluation with N-gram values 1, 2, 3, and 4, which means it calculated the precision of matched words per 1, 2, 3, and 4 words in a sentence. It can be seen that there is an increase in the BLEU score with the addition of enhancement. Comparatively, enhancement increases the BLEU-4 score by 5% better on average than standard CNN-LSTM, with the percentage increase for HE, CLAHE, and EFF being 3%, 5%, and 7%, respectively. A better result was obtained by the CNN-LSTM method with the addition EFF enhancement method. Furthermore, EFF enhancement also received better results than other methods for BLEU-2 and BLEU-3.



	High BLEU Score		Low BLEU Score	
Ground Truth	no acute cardiopulmonary abnormality	no acute cardiopulmonary process. no evidence of active tuberculosis	hyperexpanded but clear lung	no acute pulmonary abnormality demonstrated. stable cardiomegaly. prominent contour of the ascending aorta consistent with known ascending aortic aneurysm
Prediction	no evidence of acute cardiopulmonary disease	no acute cardiopulmonary disease. no evidence for metastatic disease by radiographic evaluation	stable cardiomegaly and of interstitial edema with small but increasing tortuosity of the thoracic aorta	stable cardiomegaly without evidence for acute cardiopulmonary abnormality

FIGURE 5 The example prediction of the CNN-LSTM with additional EFF.

From the test data tested on the model with EFF enhancement, the predicted description with a high BLEU score successfully predicted almost the same as the description on the ground truth on most data. Meanwhile, data with low BLEU scores have several descriptions of images that do not appear or do not match. Some examples of prediction results based on the BLEU score can be seen in Fig. 5 .

From the prediction results, the model successfully predicts according to the ground truth per word with the same definition in the description text with a high BLEU value. From the data that has prediction results in the description of more than one sentence, the model can predict similar meaning or still within the scope of the definition on ground truth with differences in word choice. Low BLEU scores were primarily found in predictions with comparisons of ground truth texts of more than one sentence. From the test data results analysis, the model's prediction results still had the exact definition as ground truth. Still, the positions and word choices were different, causing a low BLEU score, as the word occurrence evaluated BLEU.

The assumption on the influence of the low BLEU score is because in writing ground truth descriptions on the dataset, there are differences in word choice used by radiologists even though they have the same meaning, as in the descriptions of 'no acute,' 'no evidence, 'no abnormality' has the same meaning which is a healthy chest condition based on X-ray results, but due to different word selection causes the BLEU score in the model to be low.

5 | CONCLUSION

In this research, we explore the effect of applying the image enhancement method to improve predictive text coherence. From experiments using HE, CLAHE, and EFF, the implementation of the EFF method has the best visualization results for images, and the evaluation results of the predictive text are better than other methods. EFF method using exposure ratio fuse the overexpose and underexpose the image, it works better to enhance the radiograph image that is prone to uneven contrast noise during the acquisition.

The histogram graph comparison shows the distribution of the histogram after enhancement appearing more even using EFF method. ChexNet pre-trained model is used as an encoder for visual feature extraction as it was trained with the similar data we used in this study. BERT embedding is a feature extraction method in text for better contextual information. Adding an enhancement method to the model provides an improvisation of model performance, as seen from the BLEU evaluation score. Compared with CNN-LSTM without any enhancement method, HE implementation to the images got a 3% better BLEU-4 score from 0.32157 to 0.332374, and CLAHE implementation got a 5% better BLEU-4 score from 0.32157 to 0.337827. EFF implementation got a 7% better BLEU-4 score from 0.32157 to 0.34482, as seen in Fig. 2 . The best improvement was obtained in the ChexNet-LSTM model with the addition of the EFF enhancements.

For future research, using a dataset with various captions would significantly impact the predictive text. Another advanced approach could be developed to improve the predictive text's coherency. The involvement of more complex image enhancement and preprocessing can also be considered to increase model performance and feature efficiency. As for the evaluation method, it still needs to consider evaluating results with radiologists as the existing method has not been able to analyze the sentence with the same sentiment but with different wording.

CREDIT

Hilya Tsaniya: Data Curation, Writing – Original Draft, Validation, and Investigation. **Chastine Fatichah:** Conceptualization, Methodology, Formal analysis, Writing – Review & Editing, and Supervision. **Nanik Suciati:** Conceptualization, Methodology, Formal analysis, Writing – Review & Editing, and Supervision.

References

1. Chokshi FH, Hughes DR, Wang JM, Mullins ME, Hawkins CM, Duszak R. Diagnostic Radiology Resident and Fellow Workloads: A 12-Year Longitudinal Trend Analysis Using National Medicare Aggregate Claims Data. Journal of the American College of Radiology 2015 7;12(7):664–669.

2. Sutramiani NP, Suciati N, Siahaan D. MAT-AGCA: Multi Augmentation Technique on small dataset for Balinese character recognition using Convolutional Neural Network. ICT Express 2021 12;7(4):521–529.

3. Darma IWAS, Suciati N, Siahaan D. A Performance Comparison of Balinese Carving Motif Detection and Recognition using YOLOv5 and Mask R-CNN. In: Proceedings of 5th International Conference on Informatics and Computational Sciences (ICICoS) Institute of Electrical and Electronics Engineers (IEEE); 2021. p. 52–57.

4. Agus W, Darma S, Suciati N, Siahaan D. Neural Style Transfer and Geometric Transformations for Data Augmentation on Balinese Carving Recognition using MobileNet. International Journal of Intelligent Engineering and Systems 2020;13(6):349–363.

5. Sutramiani NP, Suciati N, Siahaan D. Transfer Learning on Balinese Character Recognition of Lontar Manuscript Using MobileNet. In: Proceeding of 4th International Conference on Informatics and Computational Sciences Institute of Electrical and Electronics Engineers Inc.; 2020. p. 1–5.

6. Cho K, van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) Association for Computational Linguistics; 2014. p. 1724–1734.

7. Firmawan DB, Siahaan D. Bidirectional Long Short-Term Memory for Entailment Identification in Requirement Specifications Using Information from Use Case Diagrams. In: Proceedings of 2021 International Seminar on Machine Learning, Optimization, and Data Science Institute of Electrical and Electronics Engineers Inc.; 2022. p. 331–336.

8. Zisserman KSA. Very Deep Convolutional Networks for Large-Scale Image Recognition. American Journal of Health-System Pharmacy 2018;75(6):398–406.

9. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2016-December IEEE Computer Society; 2016. p. 770–778.

10. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going Deeper with Convolutions Christian. Journal of Chemical Technology and Biotechnology 2016;91(8):2322–2330.

11. A SH, ’Ling Y, ’Liu J, ’Sreenivasan R, ’Anand S, ’Arora TR, et al. PRNA at ImageCLEF 2017 caption prediction and concept detection tasks. In: CEUR Workshop Proceedings; 2017. p. 1866–1871.

12. Ying Z, Li G, Ren Y, Wang R, Wang W. A New Image Contrast Enhancement Algorithm Using Exposure Fusion Framework. In: Proceedings of International Conference on Computer Analysis of Images and Patterns Springer Cham; 2017.p. 36–46.

13. Ahammad SH, Rajesh V, Khan PJ, Sumanth P, Sivaram G, Inthiyaz S, et al. Chexnet reimplementation for pneumonia detection using pytorch. International Journal of Pharmaceutical Research 2020;12(2):327–333.

14. Duygulu P, Barnard K, de Freitas JFG, Forsyth DA. Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary. In: Proceedings of European Conference on Computer Vision Springer Berlin, Heidelberg; 2002.p. 97–112.

15. Kiros R, Salakhutdinov R, Zemel RS. Unifying Visual-Semantic Embeddings with Multimodal Neural Language Models. CoRR 2014;p. 1–13. http://www.cs.toronto.edu/~rkiros/lstm_scnlm.html.

16. Shin HC, Roberts K, Lu L, Demner-Fushman D, Yao J, Summers RM. Learning to Read Chest X-Rays: Recurrent Neural Cascade Model for Automated Image Annotation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016 12;2016-December:2497–2506.

17. Pelka O, Friedrich CM. Keyword generation for biomedical image retrieval with recurrent neural networks. In: CEUR Workshop Proceedings; 2017. p. 1866–1871.

18. Liang S, Li X, Zhu Y, Li X, Jiang S. ISIA at the ImageCLEF 2017 image caption task. In: CEUR Workshop Proceedings; 2017. p. 1866–1871.

19. Lyndon D, Kumar A, Kim J. Neural Captioning for the ImageCLEF 2017 Medical Image Challenges. In: CEUR Workshop Proceedings; 2017. p. 1866–1871.

20. Su Y, Liu F, Rosen M. UMass at ImageCLEF caption prediction 2018 task. In: CEUR Workshop Proceedings; 2018. p. 215.

21. Zeng X, Wen L, Liu B, Qi X. Deep learning for ultrasound image caption generation based on object detection. Neurocomputing 2020 6;392:132–141.

22. Zhang Z, Xie Y, Xing F, McGough M, Yang L. MDNet: A semantically and visually interpretable medical image diagnosis network. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 2017 11;2017-January:3549–3557.

23. Naufal A, Fatichah C, Suciati N. Preprocessed Mask RCNN for Parking Space Detection in Smart Parking Systems. International Journal of Intelligent Engineering and Systems 2020 12;13(6):255–265.

24. Ayesha H, Iqbal S, Tariq M, Abrar M, Sanaullah M, Abbas I, et al. Automatic medical image interpretation: State of the art and future directions. Pattern Recognition 2021 6;114:107856.

25. Demner-Fushman D, Kohli MD, Rosenman MB, Shooshan SE, Rodriguez L, Antani S, et al. Preparing a collection of radiology examinations for distribution and retrieval. Journal of the American Medical Informatics Association 2016 3;23(2):304–310.

26. Veluchamy M, Subramani B. Image contrast and color enhancement using adaptive gamma correction and histogram equalization. Optik 2019 4;183:329–337.

27. Huang G, Liu Z, Maaten LVD, Weinberger KQ. Densely Connected Convolutional Networks. In: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) IEEE; 2017. p. 2261–2269.

28. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation 1997;9(8):1735–1780.

29. Papineni K, Roukos S, Ward T, Zhu WJ. Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL; 2002. p. 311–318.

How to cite this article: Tsaniya H., Fatichah C., Suciati N. (2022), Exposure Fusion Framework in Deep Learning-Based Radiology Report Generator, *IPTEK The Journal of Technology and Science*, *33(2):72-84*.