A Semantically Driven Model for Web Image Tagging Using Diverse Tag Selection

[1]Abhijith Roy, [2]Gerard Deepak
[1]Department of Computer Science and Engineering
[2]Department of Computer Science and Engineering
[1]National Institute of Technology, Tiruchirapalli, India
[2]Manipal Institute of Technology, Bengaluru, Manipal Institute of Higher Education, Manipal, India
roy.abhijith11@gmail.com
gerard.deepak.christuni@gmail.com

Abstract. There is a need for tagging of Multimedia content on the Web, specifically images, in addition, most of these images are from the medical domain continue to be neglected over the Web 3.0 and they need to be annotated and tagged in this era of Web 3.0 for easing the process of image retrieval and annotation based applications. This paper proposes a strategic model for Web Image Tagging of Medical Images using semantics and artificial intelligence techniques. The proposed framework facilitates Web Image Tagging with formalization of Term Sets by accepting the categories and annotations from the image dataset. A very strong learning infrastructure in terms of CNNs to classify the medical image dataset is encompassed, the presence of strategic domain relevant books helps in curating the knowledge graph, hence furthering the auxiliary knowledge and thereby fact based reasoning. The phenomenon of ontology generation also aggregates auxiliary knowledge and populates quiz knowledge which are highly relevant to the proposed framework. The Jiang-Conrath Similarity Index along with the Normalized Information Distance and the Horn's Index provides a very strong ecosystem for semantics similarity based reasoning and overall precision 95.47 %, with an F-measure of 96.66 %, lowest measure of FDR of 0.05 has been achieved by the proposed framework for tagging of Web images pertaining to Medical domain.

***Keywords*:** Web Image Tagging, Medical Images, Jiang-Conrath Similarity Index, NID, Term sets, ontology generation, Knowledge graphs

1. Introduction

In this digital age, images have become an integral part of the multimedia content that is present on the World Wide Web. Web 2.0, often called the "Social Web", was characterized by its emphasis on user-generated content, social networking, and collaborative information sharing. However , as the internet continued to evolve, it gave rise to Web 3.0, sometimes referred to as the "Semantic Web". In this new era, the focus shifted from mere data sharing to the seamless integration of data into a global knowledge framework. Despite these advancements in the Web's evolution, there remains a significant gap in the usability of multimedia content. Scattered on the web space as mere content without proper context and information, medical images in particular lack the necessary annotations that are useful for retrieval and information enrichment across the Web.

Tagging therefore is pivotal in addressing this issue. Tagging medical images adds a layer of semantics and meaning to the visual data. transforming these images from mere pixels to valuable sources of knowledge. It facilitates image retrieval, the ability to attach relevant tags to medical images would help medical professionals, researchers and patients to efficiently locate specific images within vast databases. It also enriches the relevance of medical images. By associating images with relevant metadata, the information that comes with a simple image is increased significantly.

The paper aims to put forth a model that is the best in class framework for medical image tagging in the era of Web 3.0. The paper explores the limitations, challenges that the baseline models face and what it is that the proposed framework overcomes to become the best in class framework for Web Image Tagging in the Medical Domain.

***Motivation*:** The primary motivation for the proposed framework is the need for tagging of digital content on the World Wide Web as the it is transforming to 3.0 from 2.0 and as it is in its transformative stage, there is a mandate requirement that images accessible on the Web should be annotated and tagged so that the retrieval becomes easier and contributes to the structural metadata of the Web 3.0 which is especially a requirement for medical images. These images are not catered for annotations with automatic annotation mechanism in the Web 2.0, henceforth, tagging medical images is a pressing need which is being proposed in the paper.

***Contribution*:** The primary contribution of the proposed framework is the generation of ontologies and formalization of knowledge graphs from the textbooks of general medicine, general surgery, radio diagnosis, pathology, and dermatology contributing in a high density selective domain relevant knowledge for fact based verification is a noble contribution. The presence of CNN as a strong classifier to categorize the medical image dataset which automatically selects features to sort the medical images is also quite prominent in the proposed model. The amalgamation of Normalized Information Distance, Horn's Index, and Jiang-Conrath Similarity Index at different stages in the proposed framework for semantics oriented reasoning and learning is also quite novel to the proposed model for tagging the medical images.

2. Related Works

In their influential work, Duncan et al. [1] review two decades of image processing in medical fields and discuss the persistent challenges. The paper provides valuable historical context for the development of medical image tagging methods, serving as a base for further research. Kougia et al. [2] presents a study on medical image tagging using deep learning and retrieval techniques. Their research contributes to the field by addressing the task of automated image annotation, a crucial aspect of medical image organization and retrieval. Focusing on social image tagging to capture the wide range of semantic nuances of such images. Qian et al. [3] offer insights into the use of tags to categorize medical images in various contexts. Their work sheds light on the potential of incorporating different semantic perspectives in medical image tagging. Zhou et al. [5] proposes a hybrid approach to image tagging that utilizes collaborative and machine learning. This approach has implications for the unified organization of medical images, merging collaborative and content-based tagging for a comprehensive system. Fu et al. [6] discuss recent advances in deep learning based image tagging techniques particularly in the medical domain. Their work highlights the potential of deep neural networks in automating the tagging and annotation of medical images, reducing manual effort in categorization tasks. Ma et al. [7] introduces a method to automatically generate annotations using CNN features. Their approach has applications in the medical domain, offering a way to automatically attach relevant keywords or labels to medical images, aiding in their categorization and retrieval. Anwar et al. [8] provides a critical review of the medical image analysis using convolutional neural networks (CNNs). Their paper emphasizes the importance of CNNs in automating the various tasks in medical image processing, including tagging, offering a holistic view of the field. Milletari et al. [9] introduce V-net, a fully neural network architecture capable of volumetric medical image segmentation. This segmentation capability is a crucial step in the tagging process, as it enables precise delineation of regions of interest in medical images. Lu et al. [10] discuss the application of deep learning and convolutional neural networks in medical image computing. Their work emphasizes the potential impact of these techniques on medical image tagging, as they can enhance the accuracy and efficiency of tagging and annotation tasks in the medical field.

3. Proposed System Architecture

Figure 1 depicts the proposed system architecture for the Medical Web Image Tagging which encompasses knowledge graph and semantics. The framework here is the medical image dataset is made use of, which is a categorical dataset, therefore the images already have annotations. The initial annotations along with the categories are extracted to formulate the initial Term Set. Subsequently, knowledge stack comprising several textbooks of several versions and several authors specifically in general medicine and general surgery are made use of, along with pathology and dermatology. Hence, five distinct medical domains were encompassed in order to carry out the proposed framework. Therefore, several glossaries and keywords from this array of medical practices were parsed and extracted from the indexes, glossaries and

keywords, image captions so as to formulate them into a knowledge graph using Terminal Shannon's Entropy. Using the Terminal Shannon's Entropy, the knowledge graph was curated using the information content of the entities.

The entities in the knowledge graph and the entities in the Term Set are subjected to computation of Normalized Information Distance and Horn's Index. NID is subjected to a threshold of 0.5 and Horn's Index is subjected to a threshold of 0.4 with a step deviance of 0.15. The reason for setting the threshold of NID as 0.5 and not being as stringent as 0.75 is because of the strength of NID and also a large number of initial instances to be anchored into the enriched entities and the same goes for Horn's Index.The Enriched Entities Set derived is further used as a model to generate ontologies and a detailed ontology is produced which is not restricted to any levels. The detailed ontology generated from the Term Set is achieved using OntoCollab.

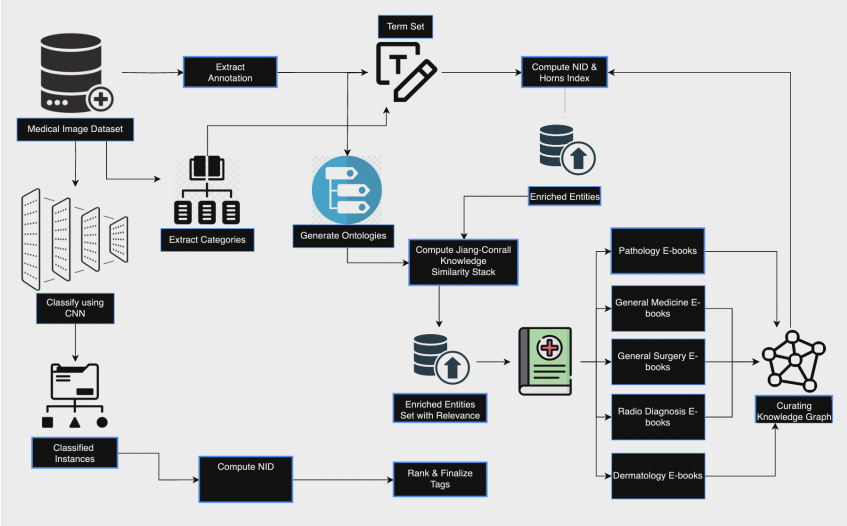

Figure 1. Proposed System Architecture for SDDS

The entities that are the result of ontology and enriched entities that are the result of computations of NID and Horn's Index are subjected to computation of Jiang-Conrath similarity measure. The Jiang-Conrath similarity measure is an ontology similarity measure which is set to a threshold of 0.60, owing to the strength of this measure. It is not made very stringent due to the anchoring of information that allows for the several layers of relevance computation.

In addition, the medical image dataset is classified using CNN as a classifier, owing to the strength and versatility of the classifier to produce accurate categorization of medical images. CNN is fed exactly with hybrid features wherein 50% of the features are set from the image content itself and the rest are extracted from the labeled or the annotations of the image dataset. The classifying instances which are produced due to the classifier are subjected to computation of NID with the Enriched Entities Set (with

Relevance Computation). It is achieved by computing NID at this juncture, NID is set to 0.70, increasing the threshold ensures high degree of relevance. The entities that are the result of the computation are ranked accordingly and are set as stack to finalization after review which is correlated to the annotations and categorizations of the medical image dataset and the tagging is finalized.

A Convolutional Neural Network (CNN) classifier is used as a deep learning model tailored for processing and classification of images and, in some cases, other structured data such as time series and text. Convolutional Neural Networks (CNNs) have found their use cases in computer vision tasks such as image categorization, object recognition, and segmentation of images. Convolutional layers apply a set of kernels on the input image to distinguish and identify attributes like edges, textures, and configurations. Pooling layers condense the spatial dimensions of the feature maps generated by the convolutional layers, helping to retain the most salient information while reducing computational complexity. Fully connected layers process the flattened feature maps, enabling the network to make predictions. CNNs use activation functions like ReLU that introduces non-linear characteristics to the model, thus having capacity to apprehend intricate data patterns.

CNN classifiers require several hyperparameters to be configured for effective training. These include the number of layers, size of kernels, type of pooling layers, number of neurons in the fully connected layers, the choice of activation functions, and the learning rate for optimisation. Hyperparameters were determined by fine-tuning on the medical image dataset through validation and experimentation. Using advanced techniques such as rotation, scaling, and horizontal flipping, can be applied to increase the model's robustness and generalization capabilities.

Training a CNN classifier involves a two step process: forward propagation and backpropagation. In the former, an input image is passed through the network, and predictions are made. The loss between the predicted and actual labels is computed, serving as an indication of the model's performance. In the backpropagation process, computing the gradient of the loss with respect to the model's parameters, and the model's weights are updated using optimisation algorithms. Process is repeated in cycles over multiple epochs until the model converges and exhibits good generalization on unseen data. CNN as a classifier is a powerful tool in enhancing image classification results with techniques as mentioned above but without regularization techniques such as dropout and batch normalization, these would succumb to overfitting.

The normalized information distance is a versatile distance metric applicable to objects of diverse types. It is based on the Kolmogorov complexity, which makes it difficult to compute but it proves useful when the objects compared are sets that are binary representations (0s or 1s). Secondly, the abstract usage of this complexity can be traced in the World Wide Web such as for names, page count statistics, and many more. NID (Normalized Information Distance) as a metric for semantics has found its use cases in machine learning, especially Image Tagging.

$$|p| = max\{K(x \mid y),\ K(y \mid x)\}$$

Formula 1. Normalized Information Distance

In this depiction of the Kolmogorov complexity, we assume the objects as finite

strings of 0s and 1s, hence, we are computing string similarity. x and y are strings and p is the most concise capable of generating x from y and vice versa. The formula has been demonstrated to hold water, even when considering logarithmic additive terms that can safely be ignored. This information distance is proven to exhibit metric properties, making it a universally applicable measure.

Horn's Index is a semantic similarity measure to calculate the similarity relation between two concepts in an ontology. It is a value in the interval (0,1), where a value closer to 1 indicates greater similarity between the two concepts.

$$HI(c_1,\ c_2) = \frac{|\,S(c_1)\ \wedge S(c_2)\,|}{|\,S(c_1)\,|}$$

Formula 2. Horn's Index

Here, c_1 and c_2 are the two concepts present in the ontology and S is the set of all subsumers of the concept. A subsumer can be defined as a concept that is more general than another concept. A subsumer can be visually as the ancestry of two terms or concepts, how much similarity exists between those two concepts.

The formula computes the semantic similarity between medical images by comparing the subsumers of the image features of the particular two images This helps in improving the accuracy of medical image tagging since it selects the most relevant tags for a given image.

The Jiang-Conrath Similarity Index is a similarity measure used in the field of natural language processing and information retrieval to compare the semantic correlation between concepts in a knowledge base. It aims to quantify the similarity of terms based on the available information content in a given ontology.

$$sim_{jcn} = \frac{1}{IC(A) + IC(B) - 2 \cdot IC(lcs(A,B))}$$

Formula 3. Jiang-Conrath Similarity Stack

1. sim_{jcn} is the Jiang-Conrath similarity between two terms A and B
2. $IC(A)$ and $IC(B)$ are the information content of terms a and b. Information content can be defined as the measure of how particular or rare a term is within a knowledge base (ontology). Terms that are more specific and less frequent are bound to have higher information content.

3. $lcs(A, B)$ is the least common subsumer, which is the maximum of the specific concepts in the ontology that is shared between the terms A and B.
4. The denominator in the formula computes the overall information content of both terms, subtracting twice the information content of their least common subsumer.

The Jiang-Conrath Similarity Index computes the semantic relatedness between two terms by comparing their information content and the specificity of their common ancestor (least common subsumer) in a knowledge ontology. In simpler terms, it quantifies how similar a term A is to term B.

In specialized domains like the Medical domain, such as is the interest of this paper, ontology based tagging is crucial. Semantic similarity measures can be used to relate image tags to concepts within an ontology. Image tagging systems combine visual content with annotation from external resources of knowledge. In such cases similarity measures such as this index can be used to bridge the gap between the image and the textual information, aiding in more decisive tagging.

4. Results and Performance Evaluation

Dataset **:** Integrative dataset comprising of Medical MNIST, Medical Images Dataset by Joe Logan (2002), and lasonas_xrist (2021) dataset for medical image processing have been strategically amalgamated together and integrated into a single large dataset by generating annotations, pre-prioritizing the entities with common annotations for the extreme annotations, scouring the Web 3.0 obtaining more images and populating the dataset for which text implementations were curated.

Implementation **:** The implementation was conducted using Python 3 as the language of choice with Google Collaboratory as the preferred IDE. Python general libraries were encompassed to conduct Natural Language Processing tasks. The dataset consisted of three verified sources in the medical domain along with the knowledge base which was updated e-books most prominent for their quiz knowledge in the distinct medical fields. Knowledge graph was curated on Neo4j using query language. CNN classifier was configured on the Keras platform.

The performance of the Semantic-Driven Medical Image Tagging model (SDDS) which is a semantically driven model for Medical Image Tagging using diverse tag selection and the results of the model is validated using various metrics such as precision, recall, accuracy, f-measure, and False Discovery Rates (FDR), where F-measure percentages yields the relevance of results, the FDR quantifies the number of false positives and thereby highlighting the error rate in the model.

From table 1, it can be inferred that the proposed SDDS outperforms comparative models, exhibiting the highest precision, recall, accuracy, f-measure percentages and the lowest value of FDR. The proposed SDDS model has achieved **95.47 %** average precision, **97.89 %** average recall, **96.68 %** accuracy, 96.66 **%** f-measure, and FDR value of **0.05**.

Model	Average Precision %	Average Recall %	Average Accuracy %	Average F-Measure %	FDR
MITD	85.82	87.08	86.45	86.44	0.15
HPIT	91.74	92.09	91.915	91.91	0.09
DCMA	93.79	94.17	93.98	93.97	0.07
Proposed SDDS	**95.47**	**97.89**	**96.68**	**96.66**	**0.05**

Table 1. Comparison of Performance of the proposed SDDS with baseline models

In order to compare the performance of the baseline models with the proposed SDDS model, it is baselined with three distinct models namely the **MITD**, **HPIT**, and **DCMA**. The MITD has 85.82 % average precision, 87.08 % average recall, 86.44 % average accuracy, 86.44 % average F-measure and FDR value of 0.15. The HPIT framework has yielded 91.74 % average precision, 92.09 % average recall, 91.915 % average precision, 91.91 % average F-measure and FDR value of 0.09. The DCMA model has yielded 93.79 % average precision, 94.17 % average recall, 93.98 % average accuracy, 93.97 % average F-measure and FDR value of 0.07. Hence, the proposed SDDS framework has achieved the highest average metrics and the lowest of value of FDR. Amongst the baseline models, the MITD is specific to the medical domain, however, HPIT and DCMA have techniques that are for distinct domains, the experimentations were conducted for the exact same dataset for the exact same domain same as for the proposed SDDS framework. The tabulations were done on the same metrics as listed above.

The reason for the proposed SDDS model to have such significant metrics compared to the baseline models is primarily because of the framework being semantically driven which operates on the medical image dataset for Web medical image tagging. The model uses a very strong infrastructure of Convolutional Neural Network Classifiers, subsequently a strong enforcement of auxiliary knowledge is seen in the model where the annotations and the categories are extracted from the medical image dataset to formulate the term set. Subsequently, the entity enrichment is achieved through general medicine books and radio diagnosis ebooks from which the knowledge graph was generated Selectively, permutations and combinations of the term set and the knowledge graph, with the use of the Horn's Index are enforced with respective step deviance methods in order to yield the enriched entities. Most importantly, the Jiang-Conrath Similarity Index, the Normalized Information Distance Index, and the Horns Index prove to be very strong and semantically driven reasoning to measures of semantic similarity.

Since the proposed framework has a strong strong infrastructure of knowledge,

learning paradigm, derivation of knowledge graph, and a very strong associated semantic reasoning semantic measures and also encompassed of strict ontologies which are generated from the dataset based term sets in order to yield the initial standard cognitive knowledge into the framework. Owing to all these facts, the proposed SDDS outperforms the rest of the baseline models.

The difference in the performance between the comparative models and the proposed model, even though the MITD model is a model trained on medical images, although it has a very strong learning infrastructure, ConceptCXN Detection was a major part of the framework. However, the model hybridized retrieval based methods with encoders as a strategy. Although the framework has a very strong deep learning model in terms of encoders in place, the semantics in the model is definitely quite weak as only term level semantics is chosen, which are concept level semantics. Strong semantic associated semantic similarity measures are missing. Most importantly, the auxiliary knowledge encompassed with several resources tied with MITD is not present and henceforth it superficially works but lacks drastically owing to the very strong deep learning model which operates only the dataset without encompassing knowledge from the Web 3.0, thereby increasing the cognitive gap between Web 3.0 and the knowledge which moves into the proposed framework. Hence, the MITD model lacks in regards to the proposed framework.

The reason why the HPIT framework model does not perform significantly well as the proposed framework is due to the fact that the HPIT is a Hyper probabilistic model for unified cognitive control space tagging. Here, in this framework, a tag image association matrix is encompassed, collaborative filtering is used which requires all the entities to be rated in order to calculate the tag to tag co-occurrence probability. The collaborative filtering which is based on non negative matrix filtering does not work well as it requires all the entities and images to be rated and image ratings from users can be biased, how a content based tagging in the model works well but does not match up to the expectation of the presence of a very strong learning framework. Apart from this, semantic knowledge encompassed in the framework is quite minimal and the semantic oriented similarity measures also doesn't work where henceforth this model doesn't compare to the proposed model.

The reason why the DCMA model lacks compared to the proposed model is due to Convolutional Tagging Multi Label Image Annotation. Here, in this framework, convolutional visual features with deep neural networks have been made to only work on a single dataset alone. Auxiliary knowledge encompassment is not present and semantic oriented reasoning is absent. Although it is a deep CNN with a strong infrastructure, the convolutional ranking which is made use of, it does not work well as it only depends on a sole and single dataset rather than encompassing knowledge and reasoning from various external data resources. Henceforth, DCMA also lacks compared to the proposed framework.

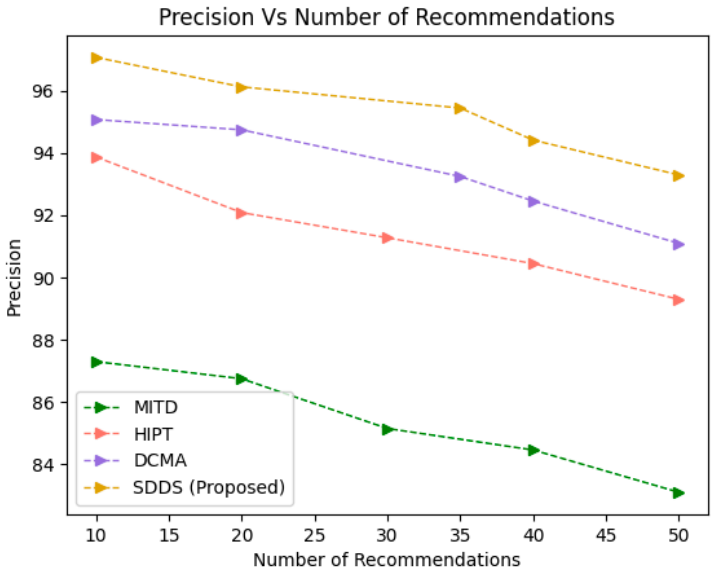

Figure 2. Comparison of Baseline Models and Proposed Model on Precision vs Number of Recommendations

The proposed framework has a very strong knowledge encompassment strategy through ontology generation from the dataset, formalization of knowledge graphs from a static knowledge stack comprising of ebooks from distinct fields of Medicine, and presence of a strong learning algorithm such as the CNN for classification, and presence of staged semantics similarity based relevance computation mechanisms such as Jiang-Conrath Similarity Index, NID, Horn's Index with the respective step deviance makes this model not only rich in knowledge, but rich in semantics which facilitates semantically oriented reasoning.

5. Conclusion

This paper proposes a strategic model for Web Image tagging focusing on images from the medical domain, in the era of Web 3.0, a very strong strategic incremental knowledge addition and derivation from the dataset starting from the annotations and categories of the dataset to formulate a term set from which ontologies are generated which are crisp instances of knowledge. Subsequently the presence of NID, Horn's index, and Jiang-Conrath Similarity Index helps in derivation of enriched entities from the knowledge graph which is generated, and helps in strategic relevance computation. General medicine ebooks, general surgery books and other domain relevant resources are used to formulate a knowledge graph which helps not only in increasing the amount of auxiliary knowledge but gives the best in-class domain based collaborative knowledge and enriches the domain which facilitates factual decision making of the framework. CNN for classifying the medical image dataset is quite apt as it is a strong deep learning model that works well for an image dataset and works on automatic feature selection as the chosen model. The Jiang-Conrath

Similarity Index also facilitates the reasoning along with NID, and Horn's Index. An overall precision of **95.47 %**, with a recall average of **97.89 %**, and accuracy measure of **96.68 %** with an F-measure of **96.66 %**, has the lowest value of FDR of **0.05** suggesting it to be the best in-class framework for Web Image Tagging for Medical Domain.

References

1. Duncan, J. S., & Ayache, N. (2000). Medical image analysis: Progress over two decades and the challenges ahead. IEEE transactions on pattern analysis and machine intelligence, 22(1), 85-106.
2. Kougia, V., Pavlopoulos, J., & Androutsopoulos, I. (2020). Medical image tagging by deep learning and retrieval. In Experimental IR Meets Multilinguality, Multimodality, and Interaction: 11th International Conference of the CLEF Association, CLEF 2020, Thessaloniki, Greece, September 22–25, 2020, Proceedings 11 (pp. 154-166). Springer International Publishing.
3. Qian, X., Hua, X. S., Tang, Y. Y., & Mei, T. (2014). Social image tagging with diverse semantics. IEEE transactions on cybernetics, 44(12), 2493-2508.
4. Zhou, Ning, William K. Cheung, Guoping Qiu, and Xiangyang Xue. "A hybrid probabilistic model for unified collaborative and content-based image tagging." IEEE Transactions on Pattern Analysis and Machine Intelligence 33, no. 7 (2010): 1281-1294.
5. Fu, Jianlong, and Yong Rui. "Advances in deep learning approaches for image tagging." APSIPA Transactions on Signal and Information Processing 6 (2017): e11.
6. Ma, Y., Liu, Y., Xie, Q., & Li, L. (2019). CNN-feature based automatic image annotation method. Multimedia Tools and Applications, 78, 3767-3780.
7. Jeya Christy, A., & Dhanalakshmi, K. (2022). Content-based image recognition and tagging by deep learning methods. Wireless Personal Communications, 123(1), 813-838.
8. Anwar, S. M., Majid, M., Qayyum, A., Awais, M., Alnowami, M., & Khan, M. K. (2018). Medical image analysis using convolutional neural networks: a review. Journal of medical systems, 42, 1-13.
9. Milletari, F., Navab, N., & Ahmadi, S. A. (2016, October). V-net: Fully convolutional neural networks for volumetric medical image segmentation. In 2016 fourth international conference on 3D vision (3DV) (pp. 565-571). Ieee.
10. Lu, Le, Yefeng Zheng, Gustavo Carneiro, and Lin Yang. "Deep learning and convolutional neural networks for medical image computing." Advances in computer vision and pattern recognition 10 (2017): 978-3.