Multimedia Tools and Applications (2021) 80:15057–15065
https://doi.org/10.1007/s11042-020-10391-w

# Deep learning based search engine for biomedical images using convolutional neural networks

Richa Mishra[1] · Surya Prakash Tripathi[2,3]

Received: 8 September 2019 / Revised: 8 October 2020 / Accepted: 22 December 2020 /
Published online: 1 February 2021

## Abstract

The development of efficient search engine queries for biomedical images, especially in case of query-mismatch is still defined as an ill-posed problem. Vector-space model is found to be useful for handling the query-mismatch issue. However, vector-space model does not consider the relational details among the keywords and biomedical image search space is not evaluated. Therefore, in this paper, we have proposed a deep learning based fusion vector-space based model. The proposed model enhances the biomedical image query similarity matching approach by fusing the vector space model and convolutional neural networks. Deep learning model is defined by converting the vector-space model to a classification model. Finally, deep learning model is trained to implement the search engine for biomedical images. Extensive experiments reveal that the proposed model achieves significant improvement over the existing models.

**Keywords** Search engine · Biomedical images · Deep learning · Websites

## 1 Introduction

Search engines are widely used in various realtime applications. Generally digital libraries and text search engines work in similar fashion. Both utilizes several indexes and utilize words for saving or retrieval of the search space [1, 25, 34]. Various searching and indexing approaches are utilized to implement the search engines. Some popular approaches are as boolean retrieval model, inverted index, etc. [11, 21, 25, 34]. However, the size of indexes become exponentially complex as number of search space increases [25]. Therefore, ranking

✉ Richa Mishra
mishraricha315@gmail.com

Surya Prakash Tripathi
tripathee_sp@yahoo.co.in

[1] Computer Science & Engineering Department, Institute of Engineering & Technology, Lucknow 226021, India

[2] Director, R. R. Institute of Modern Technology, Lucknow 226201, India

[3] Former professor Computer Science & Engineering Department, Institute of Engineering & Technology, Lucknow 226021, India

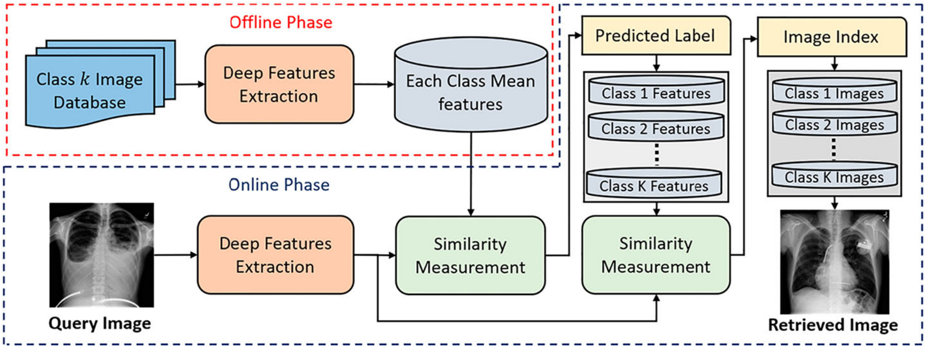

**Fig. 1** Diagrammatic representation of the biomedical search engine

of indexes is done based upon their retrieval frequency [34]. However, to evaluate the queries where similarity has significantly lesser values, is still a challenging task [10, 14, 31]. To overcome query mismatch issue, query expansion approaches have been implemented to improve the results [28]. Various alternatives or similar query words were utilized to prevent query mismatch issue [24]. Thereafter, linguistic approaches have been implemented such as latent semantic indexing [2], term-document matrix [7], word-net [20], singular value decomposition [8], etc. Latent dirichlet allocation [3, 24, 30]. However, the development of efficient search engine queries especially, in case of query biomedical image-mismatch is still defined as an ill-posed problem.

Figure 1 shows the diagrammatic representation of the biomedical search engine. It clearly shows that an efficient training model is required to build an offline biomedical image search engine. Also, during the online phase users can pass query images and obtained the respective results.

The main contributions of this paper as as:

1. A deep learning based vector-space is proposed for improving the query similarity matching for enhancing the performance of search engines especially for mismatch queries.
2. A softmax function is defined by converting the vector-space model to classification problem.
3. Finally, deep learning model is trained to implement the search engine for biomedical images.
4. Extensive experiments reveal that the proposed model outperforms the competitive models in terms of various performance metrics.

The remaining paper is as: The literature review is discussed in Section 2. Proposed model is mathematically defined in Section 3. Experimental results and discussions are presented in Section 4. Concluding remarks are presented in Section 5.

## 2 Literature review

Pinho et al. designed a novel biomedical search engine. An extensible model for biomedical images combined with an open-source picture archiving and communication model with profile-based capabilities has been utilized [22]. Long designed a novel search engine model for a supplemental health applications. Supplemental federated search engine was also designed. Performance was evaluated on federated search engine along with website usability testing results [19]. Faroo has reviewed many biomedical search engines and found that the development of biomedical search engine is still an open area of research [6].

Hochberg et al. designed biomedical search engine to diagnose diabetes. Different models such as decision tree, logistic regression, linear regression, and random forest to diagnose diabetic patients [12]. Ye et al. designed COVID-19-related query logs to develop search engines. It was significant to learn about the epidemic's influence on users' search behavior and improve search engine to tackle comparable pandemic outbreaks in the future [32]. Young et al. implemented a search engine for diagnose HIV infected patients. A negative binomial approach was designed to estimate HIV infected patients by considering a subgroup of predictor keywords recognized by lasso regression. The Google search data was integrated with existing HIV reports [33]

Fagroud et al. designed a novel internet of things (IoT) search engine. With the advancement in IoT networks and the enhancement of the number of IoT resources, searching the data of IoT, learning IoT, recognize and list of the associated resources have become a necessity, which became possible with the presence of a various kind of IoT search engines. [5]. Kopanos et al. implemented an VarSome i.e., human genomic variant search engine [17]. Doulani et al. discussed a Scopus database and google scholar search engine. The statistical population of 118 researchers who were active in social- scientific network from 29 governmental universities were utilized. t-test and pearson correlation coefficient were implemented for search engine analysis [4].

Recently many researchers have designed various machine learning models such as to classify various kind of applications [9, 13]. However, the majority of the existing models suffer from the over-fitting issue [15, 16, 29]. Therefore, in this paper, a novel fusion mdoel by using DCNN and vector space model is proposed to achieve better results.

## 3 Proposed model

In this section, initially, vector-space model is discussed. Thereafter, deep convolutional neural network (DCNN) is presented. Finally, DCNN based vector-space model is discussed.

### 3.1 Vector-space model

In this paper, we have focused on the vector-space based query similarity matching approach for improving the performance of search engines especially for biomedical image-mismatch queries.

The vector-space model defines search space and queries as group of vector indexes. Weights define the significance of biomedical image features in query $Q$ and image space $D$ [18] as:

$$Q = (N_{Q1}, N_{Q2}, ......N_{Qr}) \tag{1}$$

$$D = (N_{a1}, N_{a2}, ......N_{al}) \tag{2}$$

To defines weights for biomedical image search space vector, $lq - aDq$ [26] is utilized. In $lq - aDq$ [26], weights are computed using two factors $lq_{ak}$ i.e., frequency of word $k$ in $D_a$ and occurrence of $k$ in collected search space $(Dq_k)$. $Dq_k$ requires weight scaling.

$W$ defines total search space in $Dq_k$. Inverse biomedical image search space frequency ($aDq_k$) of $k$ is defined as:

$$aDq_k = log\frac{W}{Dq_k} \tag{3}$$

It augments $aDq$. However, it may convert $aDq$ frequent terms with low degree [25]. A composite weight is defined by integrating $Dq_k$ and $aDq_k$. Therefore, in $lQ - aDq$ weighting, the weight of $k \in D_a$ is represented as:

$$M_{ak} = lq_{ak} \times aDq_k = lq_{ak} \times logW/Dq_k \tag{4}$$

It provides significantly more weights to words having higher frequency [23, 27]. By considering $lq - aDq$, the vector-space model computes cosine similarity (cos $\theta$) among biomedical image search space and query vectors [23, 27]. cos $\theta$ defines vector details of $D_a$ and $Q$, respectively, by utilizing the dot multiplication of two vectors and also the multiplication of their respective Euclidean values. cos $\theta$ can be evaluated as:

$$cos\theta = \frac{\overrightarrow{D_a}.\overrightarrow{Q}}{|\overrightarrow{D_a} \parallel \overrightarrow{Q}|} \tag{5}$$

By using (5), dot multiplication $|\overrightarrow{D_a} \parallel \overrightarrow{Q}|$ can be evaluated as $\sum_{k-1}^{U} M_{Qk} \times M_{ak}$. Here, $M_{Q,k}$ shows weight of $k$ in query $q$. $U$ defines size of word. $|\overrightarrow{D_a} \parallel \overrightarrow{Q}|$ defines the multiplication of Euclidean values and can be evaluated as $\sum_{k-1}^{U} M_{Q,k}^2 \sum_{k-1}^{U} M_a^2k$. Integration of these variables define the similarity among $D_a$ and $Q$ as:

$$sim(D_a, Q) = \sum_{k-1}^{U} M_{Q,k} \times M_{ak}/\sqrt{\neq of.terms.in.D_a} \tag{6}$$

Equation (6) predicts the normalization normalization impact in a search engine. However, vector-space model does not consider the relational details among the keywords and biomedical image search space is not evaluated.

### 3.2 Deep learning model

In this section, deep convolutional neural network (DCNN) based vector-space model is defined. Our goal is to predict such a combination of $Q$ and $D$ which can provide more accurate results. DCNN requires various convolution filters to squeeze local features (please see Fig. 2).

Consider there is single channel which can be defined as:

$$C = [c_1, c_2, c_3, ..., c_n]\,. \tag{7}$$

where $C \in \mathbb{R}^{n\times k}$. $n$ shows the size of input biomedical image. $k$ represents the enclosed dimension of every input factor. In convolution process, a filter $\mathbf{m} \in \mathbb{R}^{lk}$ is required in implementing to successive $l$ biomedical images to bring potential features as:

$$x_i = f\left(\mathbf{m} \cdot \mathbf{c}_{i:i+l-1} + b\right), \tag{8}$$

Here, $\mathbf{c}_{i:i+l-1}$ is the integration of $c_i, ..., c_{i+l-1}$. $b \in \mathbb{R}$ is a bias. $f$ represents a non-liner activation function like $relu$. Thereafter, filter $\mathbf{m}$ move towards $\{\mathbf{c}_{1:l}, \mathbf{c}_{2:l+1}, ..., \mathbf{c}_{n-l+1:n}\}$, then following feature map can be obtained:

$$\mathbf{x} = \left[x_1, x_2, ..., x_{n-l+1}\right]. \tag{9}$$

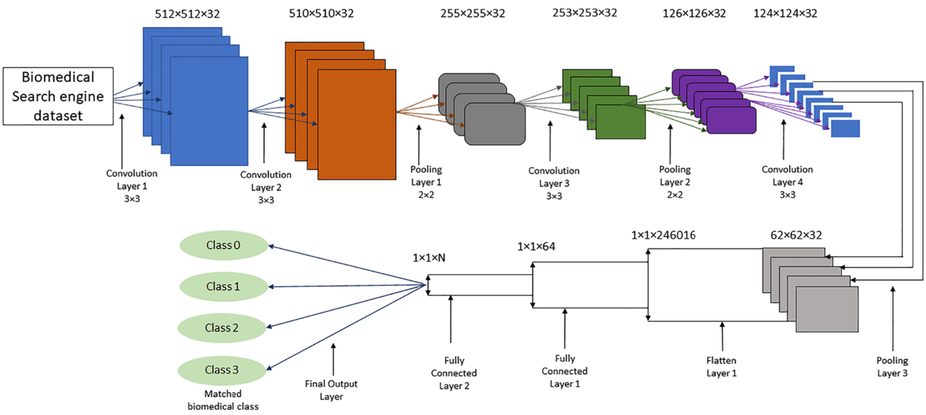

**Fig. 2** Deep learning based biomedical image search engine

Thereafter, max-pool is implemented on $\mathbf{x}$ to obtain the maximum value $\hat{x} = \max\{\mathbf{x}\}$. It defines the final feature extracted by $\mathbf{m}$. It obtains the dominated feature set of every filter. CNN computes various feature sets by using numerous filters with different sizes. The obtained feature sets contain a vector as

$$\mathbf{r} = [x_1, x_2, ..., x_s] \tag{10}$$

Here, $s$ defines the number of filters. The softmax $(s_f)$ layer is then used to compute the estimated probability distribution as:

$$y = s_f \left( W \cdot \mathbf{r} + b \right). \tag{11}$$

Consider a training data $(\mathbf{x}^i, y^i)$ in which $y^i \in \{1, 2, \cdots, c\}$ defines matched image query for search engine of $\mathbf{x}^i$ and approximated probability of DCNN is $\tilde{y_j^i} \in [0, 1]$ for

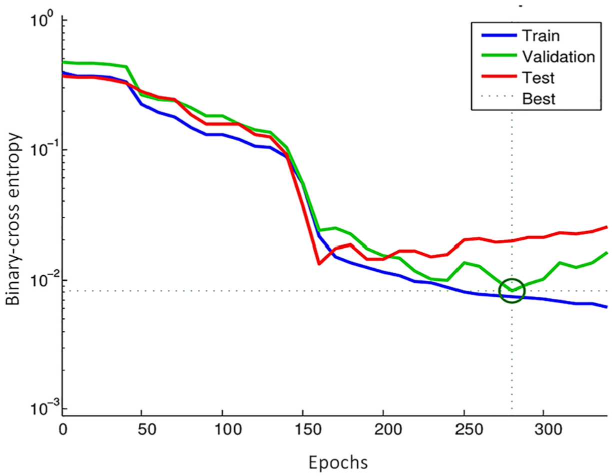

**Fig. 3** Binary-cross entropy based loss analysis of proposed model

every label $j \in \{1, 2, \cdots, c\}$. The estimated error can be computed as:

$$L(\mathbf{x}^i, y^i) = -\sum_{j=1}^{c} if\{y^i = j\}\log(\tilde{y_j^i}). \tag{12}$$

where $c$ shows the number of labels of $\mathbf{x}^i$. $if\{\dot{}\}$ define as an indicator and $if\{y^i = j\} = 1$ if $y^i = j$, $if\{y^i = j\} = 0$ otherwise. The stochastic gradient descent is employed to update the DCNN attributes and adopt Adam optimizer.

## 4 Performance analysis

The proposed model is applied to the benchmark search engine dataset. The comparison of the proposed technique is drawn with the state-of-art models such as Decision tree, Logistic regression, Support vector machine, Artificial neural network, Random forest, Naive Bayes, k nearest neighbour (k-NN), Adaboost, SVM-Random forest, CNN, and Gradient boosting. The experiments are performed on core $i7$ 3.80 GHz, 32-GB RAM, and $15M$ cache on MATLAB 2019$a$ software.

Figure 3 shows the validation, training and testing analysis of proposed model. It is found that the proposed model converges at very fast speed during the training process. At $262^{nd}$ epoch, the proposed model achieves the best training and validation results, respectively. Thus, the proposed model obtains significantly lesser binary-cross entropy values, i.e., loss during the model building process.

To evaluate the performance of the proposed model, median and degree of uncertainty values (i.e., median $\pm IQR \times 1.5$) are evaluated by repeating the experiments 50 times. We have used 65% dataset for training, 15% for validation, and 20% for testing, respectively. The fraction of training is set to be 65% because the obtained dataset is small in size. Other

**Table 1** Training analysis

| Model | Accuracy | F-measure | Sensitivity | Specificity | AUC |
|---|---|---|---|---|---|
| Decision tree | 77.64 ± 1.59 | 77.45 ± 1.72 | 77.45 ± 1.39 | 77.67 ± 1.41 | 77.59 ± 1.52 |
| Logistic regression | 77.75 ± 1.42 | 77.49 ± 1.39 | 77.66 ± 1.34 | 77.86 ± 1.21 | 77.71 ± 1.21 |
| Support vector machine | 78.86 ± 1.24 | 78.86 ± 1.32 | 78.86 ± 1.14 | 79.10 ± 0.86 | 79.11 ± 0.86 |
| Artificial neural network | 79.16 ± 0.83 | 79.12 ± 0.83 | 79.21 ± 0.77 | 79.34 ± 0.69 | 79.26 ± 0.72 |
| Random forest | 79.37 ± 0.81 | 79.42 ± 0.69 | 79.22 ± 0.59 | 79.56 ± 0.71 | 79.46 ± 0.54 |
| Naive Bayes | 79.57 ± 1.10 | 79.56 ± 1.11 | 79.62 ± 0.82 | 79.45 ± 0.68 | 79.44 ± 0.84 |
| k-NN | 79.87 ± 0.56 | 79.75 ± 0.52 | 79.57 ± 0.54 | 79.87 ± 0.85 | 79.66 ± 0.82 |
| Adaboost | 79.82 ± 0.59 | 79.82 ± 0.66 | 79.52 ± 0.58 | 79.68 ± 0.48 | 79.68 ± 0.66 |
| SVM-Random forest | 80.11 ± 0.52 | 80.32 ± 0.42 | 80.11 ± 0.52 | 80.32 ± 0.45 | 80.31 ± 0.52 |
| CNN | 80.45 ± 0.64 | 80.53 ± 0.62 | 80.34 ± 0.64 | 80.34 ± 0.64 | 80.35 ± 0.64 |
| Gradient boosting | 81.62 ± 0.45 | 82.19 ± 0.35 | 81.53 ± 0.54 | 81.57 ± 0.42 | 80.92 ± 0.48 |
| Proposed DCNN | **82.89 ± 0.36** | **82.81 ± 0.23** | **83.76 ± 0.41** | **82.72 ± 0.45** | **83.21 ± 0.42** |

Bold values indicate the higher performance

**Table 2** Testing analysis

| Model | Accuracy | F-measure | Sensitivity | Specificity | AUC |
|---|---|---|---|---|---|
| Decision tree | 76.86 ± 2.11 | 76.82 ± 2.12 | 77.13 ± 1.88 | 77.23 ± 1.71 | 77.23 ± 1.81 |
| Logistic regression | 77.21 ± 1.71 | 77.26 ± 1.72 | 77.31 ± 1.68 | 77.51 ± 1.58 | 77.35 ± 1.84 |
| Support vector machine | 77.46 ± 1.72 | 77.46 ± 1.72 | 77.54 ± 1.74 | 77.61 ± 1.53 | 77.58 ± 1.64 |
| Artificial neural network | 77.72 ± 1.52 | 77.73 ± 1.46 | 77.49 ± 1.53 | 77.82 ± 1.44 | 77.56 ± 1.29 |
| Random forest | 77.72 ± 1.34 | 77.72 ± 1.25 | 77.82 ± 1.34 | 77.77 ± 1.24 | 77.59 ± 1.29 |
| Naive Bayes | 76.22 ± 1.54 | 76.11 ± 1.39 | 77.69 ± 1.63 | 76.12 ± 1.56 | 76.12 ± 1.56 |
| k-NN | 76.33 ± 1.31 | 76.35 ± 1.35 | 76.21 ± 1.43 | 76.31 ± 1.38 | 76.28 ± 1.29 |
| Adaboost | 76.51 ± 1.12 | 76.61 ± 1.34 | 76.42 ± 1.29 | 76.53 ± 1.14 | 76.45 ± 1.24 |
| SVM-Random forest | 76.73 ± 0.72 | 76.81 ± 0.82 | 76.61 ± 0.88 | 76.63 ± 0.82 | 76.66 ± 0.82 |
| CNN | 76.71 ± 1.12 | 76.82 ± 0.81 | 76.84 ± 1.27 | 76.82 ± 1.12 | 76.82 ± 1.21 |
| Gradient boosting | 77.10 ± 0.54 | 77.23 ± 0.86 | 76.77 ± 0.86 | 76.82 ± 0.82 | 76.81 ± 0.86 |
| Proposed DCNN | **79.42 ± 0.84** | **79.42 ± 0.65** | **79.15 ± 0.77** | **79.27 ± 0.82** | **79.25 ± 0.81** |

Bold values indicate the higher performance

experiments are also considered by changing the fractions of training data. But it is found that the significant performance is found when the fraction of the training data is 65%.

To draw comparisons among the proposed and the existing models, confusion matrix-based measures are used. These measures are accuracy, specificity, sensitivity, area under curve (AUC) and f-measure.

Tables 1 and 2 depict the training and testing analysis of the proposed model for biomedical search engine dataset. Various confusion matrix based metrics like accuracy, sensitivity, specificity, f-measure, and AUC are used to compute the effectiveness of the proposed model over the existing models. From these tables, it is found that the proposed automated model provides significantly better results as compared to the existing model. As the proposed model achieves significantly better sensitivity and specificity values, therefore, a fast and efficient search engine similarity algorithm is proposed.

Table 3 shows web image search engines analysis among the proposed and the existing models. It is found that the proposed model outperforms the competitive web image serach engines.

## 5 Conclusion

From the extensive review, it has been found that the vector-space model did not consider the relational details among the biomedical contents and image search space. Therefore, a fused DCNN and vector-space based biomedical image query similarity matching approach was proposed for improving the performance of biomedical search engines. DCNN model was defined by converting the vector-space model to classification problem. Finally, biomedical image search engine was trained. Extensive experiments have been drawn by using the proposed and the competitive models for search engines. The proposed model has shown significant improvement over the existing biomedical search engines.

**Table 3** Performance analysis of the web image search engines

| Model | Accuracy | Sensitivity | Specificity |
|---|---|---|---|
| Corbis | 77.72 | 79.38 | 76.45 |
| Getty Images | 77.87 | 77.32 | 79.58 |
| Ditto | 77.09 | 76.89 | 78.45 |
| Yahoo | 77.71 | 79.09 | 78.32 |
| Picsearch | 77.88 | 76.75 | 78.32 |
| Ithanki | 77.97 | 77.54 | 79.65 |
| Web Seek | 76.58 | 76.61 | 79.06 |
| Google | 77.97 | 79.38 | 79.65 |
| Proposed DCNN | **79.23** | **79.36** | **79.07** |

Bold values indicate the higher performance

## References

1. Basavegowda HS, Dagnew G (2020) Deep learning approach for microarray cancer data classification. CAAI Trans Intell Technol 5(1):22–33
2. Bigelow JL, Edwards A, Edwards L (2016) Detecting cyberbullying using latent semantic indexing. In: Proceedings of the first international workshop on computational methods for CyberSafety, pp 11–14
3. Cimiano P, Schultz A, Sizov S, Sorg P, Staab S (2009) Explicit versus latent concept models for cross-language information retrieval. In: IJCAI, vol 9, pp 1513–1518
4. Doulani A, Shabani Z, Baradar R (2020) Information science academic members of iranian public universities sharing information resources in researchgate social scientific network: It's relation on their scientific output in scopus database and google scholar search engine. J Payavard Salamat 14(1):53–64
5. Fagroud FZ, Ben Lahmar EH, Amine M, Toumi H, El Filali S (2019) What does mean search engine for iot or iot search engine. In: Proceedings of the 4th international conference on big data and internet of things, pp 1–7
6. Faroo D (2017) Search engine optimization for medical publishing, Reconstruct Rev 7 (4)
7. Furnas GW, Deerwester S, Durnais ST, Landauer TK, Harshman RA, Streeter LA, Lochbaum KE (2017) Information retrieval using a singular value decomposition model of latent semantic structure. In: ACM SIGIR Forum, vol 51. ACM, New York, pp 90-105
8. Gao W, Guo Y, Wang K (2016) Ontology algorithm using singular value decomposition and applied in multidisciplinary. Clust Comput 19(4):2201–2210
9. Ghosh S, Shivakumara P, Roy P, Pal U, Lu T (2020) Graphology based handwritten character analysis for human behaviour identification. CAAI Trans Intell Technol 5(1):55–65
10. Gupta A, Singh D, Kaur M (2020) An efficient image encryption using non-dominated sorting genetic algorithm-iii based 4-d chaotic maps. J Ambient Intell Humaniz Comput 11(3):1309–1324
11. Gupta B, Tiwari M, Lamba SS (2019) Visibility improvement and mass segmentation of mammogram images using quantile separated histogram equalisation with local contrast enhancement. CAAI Trans Intell Technol 4(2):73–79
12. Hochberg I, Daoud D, Shehadeh N, Yom-Tov E (2019) Can internet search engine queries be used to diagnose diabetes? analysis of archival search data. Acta Diabetol 56(10):1149–1154
13. Kaur M, Singh D, Kumar V (2020) Color image encryption using minimax differential evolution-based 7d hyper-chaotic map. Appl Phys B 126(9):1–19
14. Kaur M, Singh D, Kumar V, Sun K (2020) Color image dehazing using gradient channel prior and guided l0 filter. Inf Sci 521:326–342
15. Kaur M, Singh D, Sun K, Rawat U (2020) Color image encryption using non-dominated sorting genetic algorithm with local chaotic search based 5d chaotic map. Futur Gener Comput Syst 107:333–350
16. Kaur M, Singh D, Uppal RS (2019) Parallel strength pareto evolutionary algorithm-ii based image encryption. IET Image Process 14(6):1015–1026
17. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Aguilera MA, Meyer R, Massouras A (2019) Varsome: The human genomic variant search engine. Bioinformatics 35(11):1978

18. Lee DL, Chuang H, Seamons K (1997) Document ranking and the vector-space model. IEEE Softw 14(2):67–75
19. Long BA (2017) Addressing a discovery tool's shortcomings with a supplemental health sciences-specific federated search engine. J Electron Res Med Lib 14(3-4):101–113
20. Miller GA (1995) Wordnet: A lexical database for english. Commun ACM 38(11):39–41
21. Osterland S, Weber J (2019) Analytical analysis of single-stage pressure relief valves. Int J Hydromechatron 2(1):32–53
22. Pinho E, Godinho T, Valente F, Costa C (2017) A multimodal search engine for medical imaging studies. J Digit Imag 30(1):39–48
23. Ross NC, Wolfram D (2000) End user searching on the internet: An analysis of term pair topics submitted to the excite search engine. J Am Soc Inf Sci 51(10):949–958
24. Schütze H, Hull DA, Pedersen JO (1995) A comparison of classifiers and document representations for the routing problem. In: Proceedings of the 18th annual international ACM SIGIR conference on research and development in information retrieval, pp 229–237
25. Schütze H, Manning CD, Raghavan P (2008) Introduction to information retrieval, vol 39. Cambridge University Press, Cambridge
26. Singhal A et al (2001) Modern information retrieval: A brief overview. IEEE Data Eng Bull 24(4):35–43
27. Spink A, Wolfram D, Jansen MB, Saracevic T (2001) Searching the web: The public and their queries. J Amer Soc Inform Sci Technol 52(3):226–234
28. Voorhees EM (1994) Query expansion using lexical-semantic relations. In: SIGIR'94. Springer, pp 61–69
29. Wang R, Yu H, Wang G, Zhang G, Wang W (2019) Study on the dynamic and static characteristics of gas static thrust bearing with micro-hole restrictors. Int J Hydromechatron 2(3):189–202
30. Wei X, Croft WB (2006) Lda-based document models for ad-hoc retrieval. In: Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval, pp 178–185
31. Wiens T (2019) Engine speed reduction for hydraulic machinery using predictive algorithms. Int J Hydromechatron 2(1):16–31
32. Ye Z, Mao J, Liu Y, Zhang M, Ma S (2020) Investigating covid-19-related query logs of chinese search engine users. Proc Assoc Inform Sci Technol 57(1):e424
33. Young SD, Zhang Q (2018) Using search engine big data for predicting new hiv diagnoses. PloS one 13(7):e0199527
34. Zobel J, Moffat A (2006) Inverted files for text search engines. In: ACM Comput Surv (CSUR), vol 38, pp 6–es

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.