Scalable Computing: Practice and Experience, ISSN 1895-1767, `http://www.scpe.org`
© 2024 SCPE. Volume 25, Issues 5, pp. 4360–4367, DOI 10.12694/scpe.v25i5.3085

DIAGNOSIS AND TREATMENT SYSTEM BASED ON ARTIFICIAL INTELLIGENCE AND DEEP LEARNING

XIAOXI ZHENG[*], QILI FAN[†], AND GENG WANG[‡]

Abstract. This paper designs an assisted diagnosis and treatment system based on deep learning algorithms and medical knowledge to solve the problem of poor use efficiency of massive electronic medical information. First, the disease data in the medical database is segmented to get the reverse order search table. Secondly, the similarity between the obtained clinical manifestation data and the corresponding diseases is analyzed and classified to obtain the clinical diagnosis. Then, the feedback-query method is used to analyze the weighted ratio of the original and feedback data, and the optimal fault diagnosis is carried out. The method of implicit semantic modeling is used to give the diagnosis scheme of the disease. The search method based on inference rules is introduced to realize personalized diagnosis and treatment resource recommendations to users. In this way, the specific attributes of medical resources based on individual information are effectively combined. Experiments show that the initial diagnosis recognition rate of the proposed method is 95%, the correct rate is 85%, and the recognition rate is 95% after optimization.

Key words: Deep learning; Medical knowledge base; Artificial intelligence; Electronic diagnosis and treatment; Lingo model

1. Introduction. Many patients with complicated diseases will come to first-tier cities to seek better treatment. There are many online information platforms to make the masses better understand the relevant medical information. In the past, access to medical information was mainly based on keyword-based information retrieval. However, due to the rapid development of medical data, this approach rarely meets the needs of patients. Many intelligent recommendation algorithms have recently been applied to medical data retrieval. Literature [1] provides a collaborative screening algorithm for patients with a particular disease, which can effectively solve the problem of personalized diagnosis and treatment of patients. Literature [2] proposes a medical knowledge recommendation method based on cooperation, which generates a trusted factor through the evaluation of medical services and introduces the trusted factor into the recommendation algorithm of joint filtering to realize personalized recommendations for physicians. This method not only overcomes the "information overload" of doctors but also improves the recommendation effect. Literature [3] extracts user data from user text, constructs a medical data knowledge map, and then carries out medical service resource recommendations based on collaborative filtering. Hidden semantic models such as matrix decomposition, probabilistic cryptology analysis, and implicit Dirichlet distribution effectively find hidden data. Reference [4] applies matrix decomposition to the prescription problem to discover the combination mode. However, the biggest drawback is that its parameter scale depends on the sample size, which can easily cause overfitting. Reference [5] provides the implicit semantic analysis method. They use the hidden tree method to find the hidden structure of the disease and construct the objective diagnostic criteria. The study used only disease data, not drug data, which is integral to the dialectics. Previous studies focused on the analysis of clinical manifestations. Yet, each case includes a set of symptoms and a set of medications. This project intends to use implicit-semantic modeling and case analysis methods to realize the association analysis of drug - pathogenesis. Then, the network platform provides personalized medication recommendations for patients.

2. Diagnosis and treatment based on the medical knowledge base. Reference [6] gives an architecture for assisting disease diagnosis and treatment. The system mainly includes the following parts: 1) Analysis of disease information through the characteristics of the disease segmentation operation to analyze the disease

---

[*] School of Architecture and Environmental Engineering, Zhengzhou Technical College, ZhengZhou, 450121, China (Corresponding author, `zxxdtg@163.com`)

[†] School of Automation and Internet of things, Zhengzhou Technical College, ZhengZhou, 450121, China

[‡] The sixth design and Research Institute of Mechanical Industry Co., Ltd, Zhengzhou 450100, Henan, China

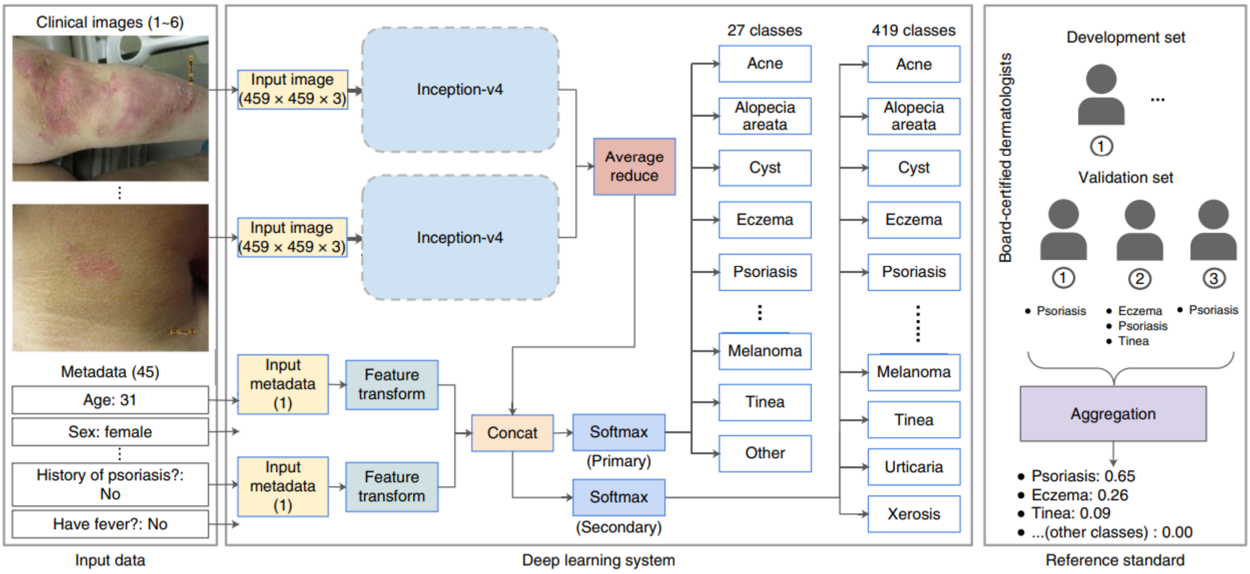

Fig. 2.1: *Structure of diagnosis and treatment system based on deep learning.*

information; 2) Disease index table an inverted index table can be constructed based on the analysis of disease information. 3) The condition assistance diagnosis module can input patient information into the system and judge the condition. 4) Diagnostic results and information feedback module displays diagnostic results. The optimal diagnosis of the disease is combined with the user's opinion. 5) The case data analysis module can obtain the detection methods related to the disease. 6) Check the suggestion form to develop the diagnosis method according to the case analysis. 7) Diagnosis The diagnostic suggestion module displays the detection patterns and diagnostic possibilities. 8) The disease association information module displays disease-related information, such as etiology, diagnosis, and treatment mode. The system analyzes the relevant information on the disease in the medical database and generates the disease search form and diagnosis mode suggestion form in the background. Secondly, based on the clinical manifestation data of the patients, the suspected disease was diagnosed, and the correlation analysis was carried out. Then, if the user has feedback on their condition, the optimal diagnosis will be made, and the details of the condition and further tests can be seen.

3. Auxiliary diagnostic technology. This section discusses specific embodiments of the assisted diagnostic techniques described in Figure 3.1.

1. Perform the same segmentation on the clinical manifestations of the input patients and compare the segmentation results with the search table. All eligible cases are entered into the potential outcome set, and the total number of these cases is called N.
2. A symptom index table was used to calculate the degree of correlation of each condition in the potential outcome set for each case and ranked according to the degree of correlation.

Use $S = \frac{1}{t}\sum(q * d)$ to express the correlation function. A similarity measurement method based on the vector space model is proposed starting from the internal product method of the vector space model. Here S represents the degree to which the patient's condition is related to the input. t is the number of types of diseases occurring in the disease, which is used to balance the matching probability of diseases with more significant symptoms in the disease database [7]. This solves the problem that the weight of the disease is too large due to the wide range of the disease. Where q is the proportion of each condition in different keywords. Treat the condition as the first possible condition. d represents the weight of each condition, again increasing the weight of the likely primary condition [8]. The correlation degree obtained by this function does not have a specific upper limit. For the convenience of calculation, the maximum and minimum values are standardized, so the correlation degree presented in the end is a relative value.

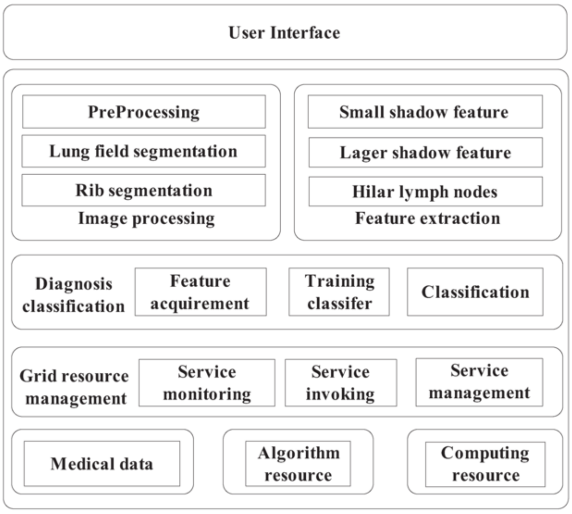

Fig. 3.1: *Deep learning-based assisted diagnosis and treatment system flow.*

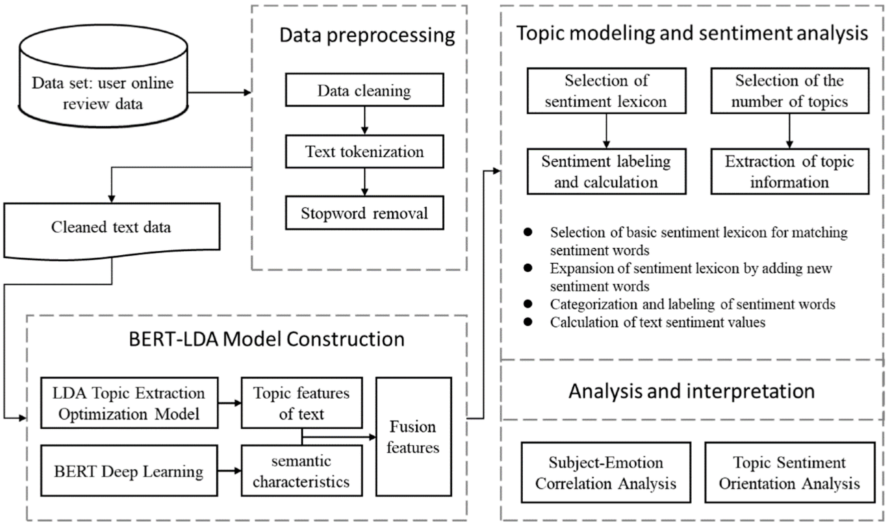

Fig. 4.1: *Linear discriminant model flow.*

4. The linear discriminant model was used to analyze medical case samples.

4.1. Introduction of Linear Authentication Mode. The implied Dirichlet distribution was first established in the multi-level Bayesian model by Blei et al. in 2003. The linear authentication pattern flow (Healthcare. MDPI, 2023, 11 (15):2142) is shown in Figure 4.1. This method was initially used in word processing and has been widely used in many data mining applications. The linear discrimination model is a typical topic model, which is tobuildsdel and disco-discovers.

A topic is a set of interrelated words that can be used to illustrate the topic. The words in the document are expressed in terms of conditional possibilities related to the topic [9]. Each topic contains words that have a high probability and are highly relevant to that topic. One word can have many different topics at the same time. Therefore, this paper proposes the ”document-topic-vocabulary” correlation based on the generation pattern. The method produces a topic with a certain probability and then a word with a certain probability. The probability of each word γ appearing in A document s can be calculated by the following formula:

$$\mid f(\gamma \mid s)=\sum_{t} f(\gamma \mid t) f(t \mid s)$$

t is the topic of this passage. In linear discrimination mode, the text contains the assignment of topics, and the topic contains the assignment of keywords [10]. Think of each prescription as a document that records the patient's condition and the medication the doctor prescribes. If symptoms and drugs are treated as "words" without distinction, they can be clustered through a linear identification pattern. Each group in the cluster contains several words, which can be a disease or a drug. Theoretically, explaining the rationality of such a cluster model is also problematic. Etiology is the subject of the model in the diagnosis problem related to conditions and both types of literature. If the two documents can be classified, establishing a link between the disease and the cause can better describe the problem.

4.2. Multiple linear discrimination model. Because of the existing methods in the aspect of data analysis of the connotation of the project plans to build a more linear differential model (figure 4.2) and reveal the whole process of production (PeerJ Computer Science, 2023, 9: e1016), the case will be regarded as files, including disease vocabulary r and γ two different types of vocabulary [11]. Given the pathogenesis, the distribution of these two types of words is independent and determined by the subject of the prescription, that is, the pathogenesis. c is a twovariable observation. There are only two values for the font: SYMPTOM or HERB. If it is $c=SYMPTOM$, the resulting word is the symptom word r. If it is $c=HERB$, the resulting word is the medical word γ. This project proposes a linear identification method based on multiple connotations. Case and drug words have the same topic, and there is a certain correlation between the two words under a specific topic. The spatial distribution of the problem c is obtained by using polynomial distribution characteristics. The post-disease word can be obtained according to the polynomial distribution lattice consistent with the problem. The medical word γ is derived from a polynomial distribution that agrees with the topic. It is the polynomial distribution lattice ζ and δ is the Dirichlet priori of χ.

After the distribution parameter $\varepsilon, \varphi, \delta$ is known, the joint probability of the "prescription - disease" distribution β, "disease - disease" distribution ζ, and "disease - drug" distribution χ is obtained as follows:

$$\begin{aligned}
& f(r, \gamma, c, c, \beta, \zeta, \chi \mid \varepsilon, \varphi, \delta)= \\
& f(\zeta \mid \varphi) f(\chi \mid \delta) f(\beta \mid \varepsilon) f(r, \gamma, c, c \mid \zeta, \chi, \beta)= \\
& f(\zeta \mid \varphi) f(\chi \mid \delta) f(\beta \mid \varepsilon) f(r, \gamma, c, c, \zeta, \chi) \\
& f(c \mid c, \beta) f(c)
\end{aligned}$$

The Gibbs sampling method is used to identify the parameters in the system [12]. In the multi-connotation linear identification model, the possibility of disease $f\left(r_i \mid r_i^{\prime}\right)$ and drug $f\left(\gamma_i \mid \gamma_i^{\prime}\right)$ belonging to topic $c=\{1,2, \text{ L}, K\}$ was calculated for the drugs for disease $r=\left(r_1, r_2, \text{ L}, r_S\right)$ and disease $\lambda=\left(\lambda_1, \lambda_2, \text{ L}, \lambda_W\right)$ in the observed samples respectively, and iteration was carried out for each stage.

5. Data online assisted diagnosis and treatment algorithm. A multi-connotation linear identification method obtained the corresponding relationship between disease and drug. A reasonable medication plan can be obtained using the established mathematical model for clinical diagnosis and then taking the clinical manifestations of patients as input. First, the "pathogenesis" is extracted based on the known disease characteristics, and then the drug is administered according to the cause [13]. The second method establishes a "disease-drug" matrix to obtain the corresponding relationship between each disease and each drug. In order of the conditions provided, get the corresponding medication recommendations. Therefore, this project intends to adopt a hybrid model to merge the two algorithms and improve the recommendation accuracy.

5.1. Pathology-based drug recommendation methods. Assuming that the clinical manifestation of the patient is $r=\left(r_1, r_2, \text{ L}\right)$, the "disease-topic" distribution ζ, "drug-topic" distribution χ and prior parameters obtained by the multi-connotation linear identification model, then the etiological drug recommendation method can be expressed as:

Enter: A list of the patient's symptoms.

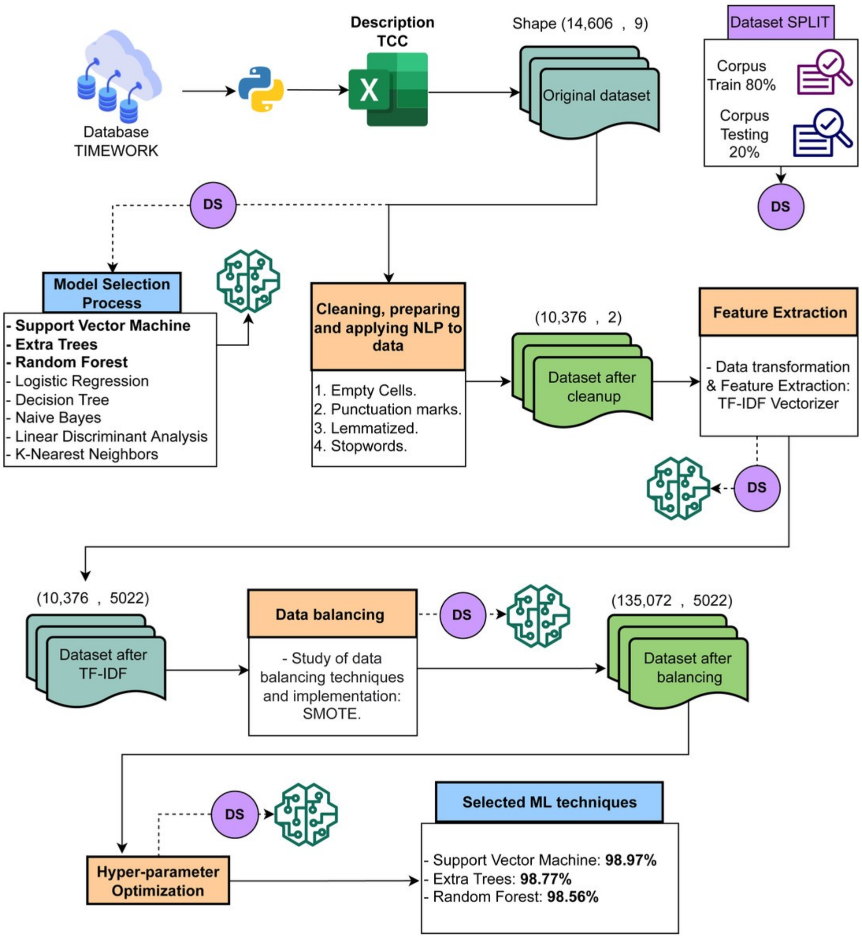

Fig. 4.2: *Multi-content linear discriminant probability graph model.*

Output: According to the patient W disease, medication is recommended.

1. Random initialization: Each current disease r will be randomly assigned a topic number c.
2. According to the sampling equation, the current file is tested again. For each disease r, the topic is sampled again.
3. The above treatment is repeated until the sampling convergence is achieved.
4. This paper determines the pathogenesis of the disease based on the distribution of the obtained topics. Top N drugs related to this disease are listed as recommendations [14]. First, the clinical manifestations of patients are regarded as the new document topic, then the disease information in the new literature is obtained by inference algorithm, and then the drug related to this disease is selected according to probability. The sampling equation to be used in the second step of the method is:

$$
\begin{aligned}
& f\left(c_{i_r}=t \mid c=SYMPTOM, c_{-i_r}, r, \gamma\right) \\
& \propto \frac{W_{r_m^{(t)},-i_r}+W_{\gamma_m^{(t)}}+\varepsilon_t}{\sum_{t=1}^K\left(W_{r_m^{(t)},-i_r}+W_{\gamma_m^{(t)}}+\varepsilon_t\right) \sum_{f=1}^S\left(W_{r_t^{(f)},\neg i_r}+\varphi_f\right)}
\end{aligned}
$$

$W_{r_m}^{(t)}$ and $W_{\gamma_m}^{(t)}$ represent the symptoms section in document m. The number of diseases listed in the title t and the number of drugs used. $W_{r_t}^{(f)}$ is the number of disease words with an average f number in the topic t.

5.2. Drug recommendation methods based on clinical manifestations. The process of drug recommendation based on disease is shown in Figure 5.1, with a set of conditions and weights of patients as their inputs. It can be expressed as an ndimensional vector. The default value is 1 ; otherwise, it is 0 . When the medical record rights are entered, the corresponding value is the input value [15]. This method expresses the correspondence between each disease and each drug by establishing a "disease-drug" matrix. Finally, according to the clinical manifestations of the patients, the most closely related to the clinical manifestations of the drug regimen.

Input: A set of symptoms for the patient and the proportion of each symptom.

Output: According to the patient W disease, medication is recommended.

1. Establish an "etiology - drug use" model. The information in column j, row i, of the matrix is $f(\gamma_j \mid r_i)$, which is the likelihood of taking the drug j when disease i occurs.
2. The rank value of each drug was calculated according to the clinical manifestations of patients and corresponding weights.
3. The drugs were ranked according to the rank, and the top W ranked drugs were recommended to patients. The formula used to construct the first step of the matrix is:

$$f(\gamma_j \mid r_i) = \sum_{c=1}^{K} f(\gamma_j \mid c) \cdot f(c \mid r_i)$$

c represents the topic in multinomial linear discrimination mode. K is the number of topics. The expression of $f(\gamma_j \mid c)$ as the parameter $\chi, f(c \mid r_i)$ in the multi-connotation linear discrimination model is as follows:

$$f(c \mid r_i) = \frac{f(r_i, c)}{f(r_i)} = \frac{f(r_i \mid c) \cdot f(c)}{f(r_i)} \propto f(r_i \mid c) \cdot f(c)$$

$f(r_i \mid c)$ is the parameter ζ in the multi-content $\mid$ model. $f(c)$ is the subject of the disease, which is obtained by the model derivation process. The grade value of each drug γ_j is calculated using the following formula:

$$\operatorname{rank}(\gamma_j) = \sum_{i=1}^{S} f(\gamma_j \mid r_i) \cdot \operatorname{weight}(r_i)$$

weight (r_i) is a weighting of symptoms r_i provided by the user.

By default, conditions that have been typed have a weight of 1, and conditions that have not been typed weight of 0.

The first medication method analyzes the pathogenesis based on the disease and prescribes reasonable medication according to the cause. The latter is to give the corresponding treatment plan for the patient's condition [16]. The focus of the two types of treatment is different. In treatment, the doctor will start with the cause and then prescribe the corresponding drug to the patient. This paper intends to design a hybrid recommendation algorithm. Take the intersection of the two as the final recommendation.

6. Experimental results and analysis.

6.1. Data Sets. 900 cases were obtained by eliminating invalid samples based on the information of 1000 patients in the modern pharmacological database [17]. Instead of measuring each case, the paper classifies each case into a class or category of diseases.

6.2. Baseline Method. The benchmark method chosen in this article is the most popular drug recommendation method. Recommend the most popular products to customers [18]. The drug recommendation questions were all for cases of lung cancer patients, so several commonly used drugs were found based on the analysis of these cases.

Table 6.1: *Statistics of modern medical records.*

Project name	**Quantity**
Caseload	900
Number of disease species	85
Total number of medicinal materials	123

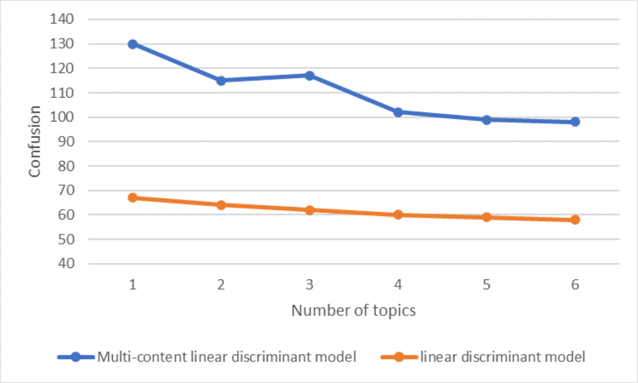

Fig. 6.1: *Results of confusion degree experiment.*

6.3. Degree of confusion of multiple linear discrimination modes. The results of multi-content linear discrimination are compared with those of traditional linear discrimination. It can be seen from Figure 6.1 that the effect of the multi-content linear discrimination model is significantly better than that of traditional linear discrimination [19]. In addition, the degree of ambiguity decreases with the increase of the number of categories Z, which indicates that the degree of ambiguity tends to converge, which is consistent with the theoretical argument. Increasing Z will not improve the efficacy after Z is large enough to cover all the hidden causes.

7. Conclusion. The improved linear identification model was used to analyze clinical cases and find hidden causes. Find out the internal relationship between recessive cause, syndrome and medication. Two methods of drug recommendation based on symptom were designed using the correlation between symptom, pathogenesis and drug. The algorithm's effectiveness is verified by testing the modern medical case base.

REFERENCES

[1] Tran, K. A., Kondrashova, O., Bradley, A., Williams, E. D., Pearson, J. V., & Waddell, N. (2021). Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Medicine, 13(1), 1-17.

[2] Goecks, J., Jalili, V., Heiser, L. M., & Gray, J. W. (2020). How machine learning will transform biomedicine. Cell, 181(1), 92-101.

[3] Yu, K., Tan, L., Lin, L., Cheng, X., Yi, Z., & Sato, T. (2021). Deep-learning-empowered breast cancer auxiliary diagnosis for 5GB remote E-health. IEEE Wireless Communications, 28(3), 54-61.

[4] Ellahham, S. (2020). Artificial intelligence: the future for diabetes care. The American journal of medicine, 133(8), 895-900.

[5] Rahman, A., Hossain, M. S., Alrajeh, N. A., & Alsolami, F. (2020). Adversarial examples—Security threats to COVID-19 deep learning systems in medical IoT devices. IEEE Internet of Things Journal, 8(12), 9603-9610.

[6] Yuan Liu, Ayush Jain, Clara Eng, David H. Way, Kang Lee, Peggy Bui, Kimberly Kanada, Guilherme de Oliveira Marinho, Jessica Gallegos, Sara Gabriele, Vishakha Gupta, Nalini Singh, Vivek Natarajan, Rainer Hofmann-Wellenhof, Greg S. Corrado, Lily H. Peng, Dale R. Webster, Dennis Ai, Susan J. Huang, Yun Liu, R. Carter Dunn & David Coz. A deep learning system for differential diagnosis of skin diseases. Nature medicine, 2020, 26(6): 900-908.

[7] Chang, V., Bailey, J., Xu, Q. A., & Sun, Z. (2023). Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. Neural Computing and Applications, 35(22), 16157-16173.

[8] Char, D. S., Abràmoff, M. D., & Feudtner, C. (2020). Identifying ethical considerations for machine learning healthcare applications. The American Journal of Bioethics, 20(11), 7-17.

[9] Quer, G., Arnaout, R., Henne, M., & Arnaout, R. (2021). Machine learning and the future of cardiovascular care: JACC state-of-the-art review. Journal of the American College of Cardiology, 77(3), 300-313.

[10] Alimadadi, A., Aryal, S., Manandhar, I., Munroe, P. B., Joe, B., & Cheng, X. (2020). Artificial intelligence and machine learning to fight COVID-19. Physiological genomics, 52(4), 200-202.

[11] Echle, A., Rindtorff, N. T., Brinker, T. J., Luedde, T., Pearson, A. T., & Kather, J. N. (2021). Deep learning in cancer pathology: a new generation of clinical biomarkers. British journal of cancer, 124(4), 686-696.

[12] Karar, M. E., Alsunaydi, F., Albusaymi, S., & Alotaibi, S. (2021). A new mobile application of agricultural pests recognition using deep learning in cloud computing system. Alexandria Engineering Journal, 60(5), 4423-4432.

[13] Khan, S., Barve, K. H., & Kumar, M. S. (2020). Recent advancements in pathogenesis, diagnostics and treatment of Alzheimer's disease. Current neuropharmacology, 18(11), 1106-1125.

[14] Chen, C., Liu, B., Wan, S., Qiao, P., & Pei, Q. (2020). An edge traffic flow detection scheme based on deep learning in an intelligent transportation system. IEEE Transactions on Intelligent Transportati on Systems, 22(3), 1840-1852.

[15] Muhammad, K., Khan, S., Del Ser, J., & De Albuquerque, V. H. C. (2020). Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey. IEEE Transactions on Neural Networks and Learning Systems, 32(2), 507-522.

[16] Oh, Y., Park, S., & Ye, J. C. (2020). Deep learning COVID-19 features on CXR using limited training data sets. IEEE transactions on medical imaging, 39(8), 2688-2700.

[17] Saood, A., & Hatem, I. (2021). COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet. BMC Medical Imaging, 21(1), 1-10.

[18] Alsharif, W., & Qurashi, A. (2021). Effectiveness of COVID-19 diagnosis and management tools: A review. Radiography, 27(2), 682-687.

[19] Li, T., Zhao, Z., Sun, C., Cheng, L., Chen, X., Yan, R., & Gao, R. X. (2021). WaveletKernelNet: An interpretable deep neural network for industrial intelligent diagnosis. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 52(4), 2302-2312.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: Dec 27, 2023
Accepted: Jan 14, 2024