Research on Feature Fusion Methods for Multimodal Medical Data

Zhaogang Xu, Xi Yang, Yu Jin, and Shuyu Chen[✉]

Chongqing University, Chongqing 400044, China
610922554@qq.com

Abstract. With the rapid development of artificial intelligence, knowledge graph, image processing, etc. have been widely used, and smart medical care, as a major application scenario of artificial intelligence, has received a lot of attention. Traditional diagnostic methods have problems such as low accuracy and low efficiency, and the research and application of knowledge graph and image classification in the field of dermatology are also in the initial stage, but text-based knowledge graph technology and image-based image classification technology have developed very maturely. Considering that various current image classification algorithms extract features, feature calculation, and model matching from images, they do not consider obtaining information such as features or relationships that are not in images from text data to participate in image classification tasks. In this paper, the optimized hierarchical perception model H-HAKE based on hierarchical perception model KGE-HAKE calculates selector parameters by improving the hierarchical perception model to add category dimension to the TransE coordinate system, divide more image features and entities with the same attribute into the same level, increase the number of links between image and map entities, and produce better data coverage effect. Aiming at the image classification task, this paper proposes a game tree model to optimize the classification results, including calculating the confidence degree based on the map, the aggregation degree of the classification results, the inference value of the entities in the domain, etc., and comprehensively designing the fusion mode of knowledge graph and image classification algorithm KG-based CNN in scenarios such as multi-map input and feature pre-extraction. The mode is effective enough to enable the image classification task to utilize multimodal data, and the effectiveness is verified by multi-scenario and data ablation experiments on the public data collection.

Keywords: KEG-HAKE · Entity Extraction · Knowledge Reasoning · Image Classification

1 Introduction

1.1 A Subsection Sample

At present, the number of skin disease patients worldwide has reached about 420 million, of which the number of skin disease patients in China alone has exceeded 150 million, and the incidence rate has reached $40\%-70\%$. Traditional diagnostic methods often rely

on the experience of doctors and visual observation, which has problems such as low diagnostic accuracy and susceptibility to subjective factors. Computer assisted diagnosis (CAD) technology has made significant progress in this social context. However, current image classification methods have some limitations, such as the lack of background features in some images, difficulty in extracting fuzzy features from images, and the fact that patients often lack professional medical terminology, they may omit or provide incomplete symptom descriptions, increasing the difficulty of disease diagnosis. Therefore, the key to handling the preliminary diagnosis of skin diseases is to confirm the disease range of patients according to the description of less symptom information, use the domain knowledge atlas to carry out and complete the missing information, and combine the text information with the patient's disease image to identify.

In China, research on disease knowledge maps and skin image diagnosis is also gradually emerging. In recent years, more and more research teams have begun to explore the use of artificial intelligence technologies such as deep learning for automated diagnosis of diseases. For example, Professor Gong Lejun's team of Southeast University uses knowledge maps to model long non coding RNA (LncRNA), describes data and corresponding relationships through Resource Description Framework and Web Ontology Language, and studies the relationship between genes and diseases [1]. Li Guojing, Xia Qiuting, and others used attention mechanism and multi model fusion based on ResNet for fundus retinal image assisted diagnosis [2]. Liu Zhaorui, Zhang Yi, and others from Peking University Union Medical College conducted a convolutional neural network (CNN) binary classification model based on dermatoscopy images for the differential diagnosis of mycosis fungoides (MF) and inflammatory skin diseases under dermatoscopy, with an accuracy of 75.02% [3].

Internationally, there has been some relevant research and practice on the construction of skin image diagnosis systems based on the knowledge graph of skin diseases. For example, Tao Meng and Lin Lin proposed a subspace projection model (C-RSPM) at the IEEE multimedia seminar to divide cell images into 25 blocks to reduce computational spatial complexity, and to use multimodal data for voting on image blocks to improve accuracy [4]. At the ACM International Multimedia Conference, A Znaidia and A Shabou proposed the concept hierarchy and the data collected from the photo sharing platform and images to complete the photo multi label marking work. Through the integration of two text and one image classification mode, computational complexity was reduced [3]. At the 2020 IEEE/CFF Symposium on Computer Vision and Pattern Recognition, E Raisi and SH Bach proposed conducting neural network training to improve model generalization ability while simultaneously training two related tasks. One is the original task (target), and the other is the auxiliary task (source). The auxiliary task is completed using a knowledge graph, reducing the error rate by 2.1% [5].

Overall, with the continuous development and application of artificial intelligence technology, skin image diagnosis systems have been widely studied and applied internationally and domestically. There is a trend towards multimodal data fusion. However, the image diagnosis system based on dermatology knowledge map has not been formally developed due to the lack of domain knowledge base and the construction model of multimodal data in image classification. However, the research on this system is expected to

further improve the diagnostic accuracy of skin diseases and promote the true digitization and intelligence of clinical practice. Therefore, it is necessary to explore new ways to integrate knowledge graphs into image classification modes.

2 Background Knowledge

2.1 TransR

The knowledge graph completion technology is an important technology for constructing knowledge graphs. The methods for inferring new entities and relationships based on the entity relationships of knowledge graphs can be summarized as follows: translation models, bilinear models, and neural network models.

Knowledge representation is the projection of the real world into the virtual world, just like computer compilation, transforming human knowledge into a form that computers can understand and easily calculate. This form can be used to simulate human reasoning about the world and then handle difficult problems in the field of artificial intelligence. Knowledge representation learning: The process of low-level embedded knowledge representation of entities and relationships in a knowledge graph.

Common Translation Distance Model Knowledge Representation Learning Methods:
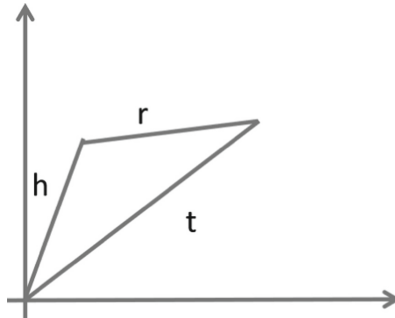
2.2 TransE

(Fig. 1)

Fig. 1. TransE.

TransE proposed a translation based knowledge graph embedding model that can capture the phenomenon of translation variance invariance in multiple relationship graphs [6]. The facts in the knowledge graph are represented by triples, and the idea of TransE algorithm is very simple, which has the same translation invariance characteristics as another text computing model, Word2vec.

TransE's loss function:

L\left(y,y^\prime\right) = \max\{\left(0,margin-y + y^\prime\right)\}

The construction process is to minimize the value of the loss function. Due to having only one entity space, the disadvantage of the model is that when dealing with multi entity problems, the calculation of the distance function may result in incorrect relationship links due to the fact that the entities are the same. For example, there are two types of knowledge, (Ogawa, get, girlfriend) and (Ogawa, get, dermatosis). After training, in a one-to-many situation, the "girlfriend" entity vector will be very close to the "psoriasis" entity vector. But in reality, they do not have such similarity.

The TransE model defines vectors in semantic space as entities and relationships, but this method can cause problems when conducting relationship inference. The calculation distances of the same entity in space are very close, but the relationships can be very different because there are more than one type of entity relationship, and distance based calculations cannot obtain such a distinction. For example, the relationship between (anti-inflammatory drugs, including, cephalosporins) is inclusive, while the relationship between (Ogawa, Allergy, cephalosporins) is allergic. These two relationships are very different (Fig. 2).
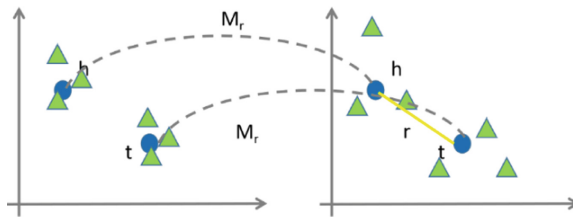

Fig. 2. TransR.

Due to the extremely close distance relationship calculated by the same entity in the entity space, TransR technology allows the same entity to be projected into two different spaces, that is, the entity space and multiple relationship spaces (relationship specific entity spaces) to model entities and relationships, and then perform transformation operations to solve the proposed problem. In this way, all entity relationship transformations become transformations in the corresponding space. Therefore, this method is named TransR [7]. It can be understood that the entity relationship in the original space is close because the entities are the same, and the relationships between adjacent entities are also close. However, if we switch to different spaces, the relationships between adjacent entities are different, and differences can be reflected.

ResNet

ResNet was the champion of the 2015 ImageNet competition, reducing the error rate of image classification recognition to 3.6%, which even exceeded the accuracy of normal human eye recognition.

The model also follows the development direction of the VGG model by adding more network layers. Theoretically, the more layers the model has, the more parameters the original layer can learn. The new layer can learn more parameters, which increases the solution space. However, in reality, the network error will become larger. ResNet proposed a solution here, based on its basic idea.
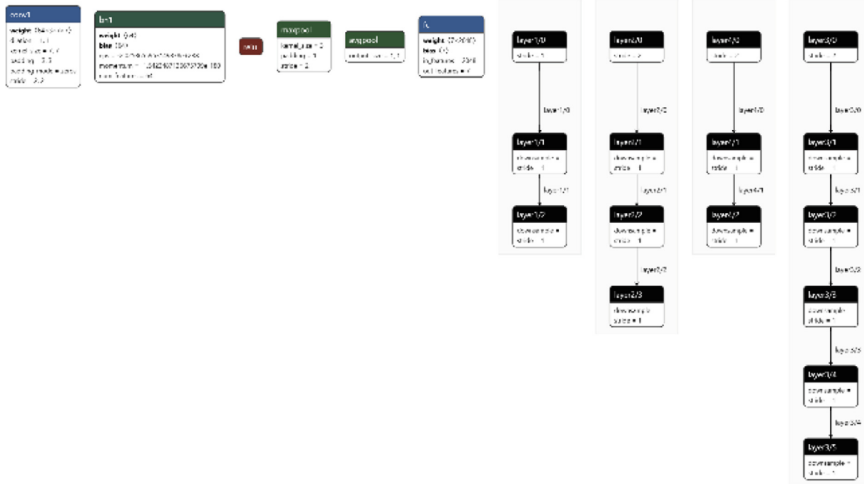
Fig. 3. ResNet network structure.

Not learning complete data, only learning residual terms means learning the Y sample residual space, not the X sample space. If you want to learn the same network, you can make X = Y, and by learning Y-X, you can deepen the training layers of the model and reduce model error [8].

3   Knowledge Graph Construction

Construction of a knowledge map of skin diseases At present, there are no researchers building a specialized Chinese knowledge map of skin diseases. Therefore, we use crawler technology to collect data on publicly available skin disease descriptions, diagnosis, and treatment information on internet medical websites. This article first evaluates internet medical websites based on the World Health Organization's disease classification standard documents, and selects data sources with rich disease symptom descriptions and hierarchical classification information, The selected data sources are the Medical Search and Medicine Website (www.XYWY.com) and the 99 Health Website (www.99.com.cn).

The following is the collected text data of skin diseases. Currently, there are about 1000 known types of skin diseases in the world, with over 200 common diseases. The two atlas datasets used in this article are 271 common diseases and 789 universal diseases. This article will store the collected text data in JSON format, classify and label it according to name, symptoms, drugs, food, etc. Then, use Re to clean the data and delete the segmented text parts, such as numbers and symbols. Use Jieba to segment and label the text data, and use Word2vec for feature processing. The following figure shows the collected text information .

Then we construct the text information into a knowledge graph of dermatology, where we use Eno4j for storage and representation to facilitate observation of changes in entity relationships during subsequent operations. Neo4j is a high-performance NoSQL

Fig. 4. Crawled Text Information.

graphical database, which stores structured data on the network (called graph from a mathematical point of view) rather than in tables. Neo4j can also be seen as a high-performance graph engine that has all the characteristics of a mature database. The construction process is to define the Eno4j connection; Define entity nodes, such as disease symptom overview drugs, etc.; Define entity relationships such as disease examination items, disease symptoms, disease complications, disease departments, and then create other nodes with the disease as the center node, and finally create relationships. The following figure shows the constructed knowledge graph.

Fig. 5. Eno4j Knowledge Map.

In the practical application process of knowledge graph, we found that there are many hidden entities and entity relationships that have not been created. In this section, after comparing the characteristics of translation models, bilinear models, and neural network models, we selected a translation model, also known as the translation vector model. Because the goal of this study is to enable image classification algorithms to use textual

information from patients, whose expressions are often fuzzy and incomplete, and the output of image classification generally corresponds to disease entities in the knowledge graph (the use of links between image classification and symptom information will be mentioned later), this article hopes to extract information from patients that can limit the scope of image classification, Referring to the selection of departments during the hospital consultation process and the consultation process, traditional Chinese medicine students first determine the approximate range and then determine the type of disease based on specific medical examination data. The process of constructing a domain limited knowledge graph is equivalent to the process of consultation, while image classification is a simulation of the analysis of physical and chemical test results [10] (Fig. 6).

Fig. 6. Skin diseases of different parts and forms.

The research in this article combines text information such as entity shape features, part features, symptom features, and image information as a multimodal data source to assist in disease image classification. Even based on the construction of entity recognition within the field, it is supported by medical theory. According to the analysis of the content of the commonly used teaching materials of Peking University Medical Department and Medical College, the classification of skin diseases in the medical diagnosis textbooks of skin diseases starts from the naming of morphology, such as psoriasis, and the gradual transition from erythema multiforme to the classification of etiology and pathogenesis, such as infectious diseases, viral fungal dermatitis, autoimmune diseases, eczema, drug rash, genetic skin diseases and other ichthyosis keratosis [9]. The diagnosis of the disease as shown in Figs. 3, 4 and Figs. 3, 5 can be based on the morphology of the skin lesion, site of infection, pathogen, dietary habits, drug reactions, etc., and there is a diagnostic sequence that gradually narrows the scope from the surface to the inside [10].

Combining computer knowledge graph technology to model medical information, the translation vector model TransE constructed by disease entities has modulus and phase information. Since the output of image classification generally corresponds to disease entities of knowledge graph types, this article hopes to find disease entities with similar positions at the same level, and modulus and phase information can help construct knowledge graphs with hierarchical relationships. Therefore, this article adopts Hierarchical Perceived Knowledge Graph Embedding (HAKE) as the basic model for model completion. Two parameters, modulus and phase information, are constructed to construct a polar coordinate system, mapping entities to the polar coordinate system.

The concentric circles in the polar coordinate system can naturally reflect the hierarchical characteristics. The angular coordinates in the coordinate system can distinguish different entities at the same level (Fig. 7).
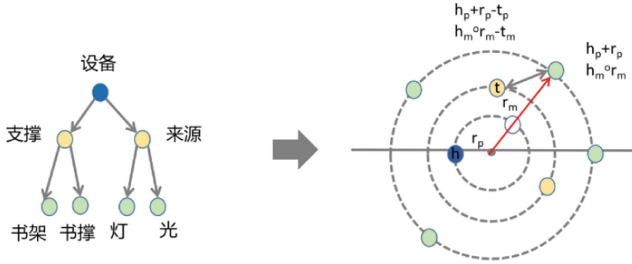

Fig. 7. KEG-HAKE Schematic Diagram.

The above figure is a schematic diagram of KEG-HAKE. The modulus parameter m in the figure represents the modeling of entities with the same hierarchical structure, while the phase information p represents the modeling of entities with different hierarchical structures. m and p can be used to represent all entities in the entity space [11].

When performing entity completion, calculate the distance between two entities, represented by the following formula.

The research purpose of this article in terms of knowledge graph is to provide domain limited knowledge graphs for image classification, and to appropriately extend knowledge in the graph when there are keywords related to entity relationships input by patient knowledge Q&A. As most of the entities obtained from Q&A activities are located at the edge of the graph, in order to extend to more entities that can be utilized by image classification, the KEG-HAKE knowledge embedding method needs to be used for entity embedding, Construct a hierarchical perceptual knowledge graph, so that we can control the scope of knowledge extension according to the hierarchy. If we can further control the direction of knowledge extension, we can solve the problem of knowledge boundaries that are not interconnected. For example, in knowledge Q&A tasks such as (Ogawa, infection, tinea manus), (Ogawa, infection, tinea cruris), (dermatosis, including, tinea manus), and (dermatosis, including, tinea cruris), we can expand knowledge to the center and draw the conclusion that Ogawa has dermatosis. In addition, for image classification tasks optimization, Since the classifier rating in the following text needs to be based on whether it is within the domain as a parameter, centripetal expansion can reduce errors. However, KEG-HAKE cannot divide entities with the same attribute into the same level. Therefore, we optimize based on the KEG-HAKE model to add hierarchical features to the graph [12]. Through the analysis of the HAKE model paper, it was found that the key to the hierarchical perception characteristics of the model lies in the fact that it is based on the TransE model as the basic model, and then adds the concepts of modulus and phase information. Analyzing the specific code of the model, it was found that the HAKE model and the reference model, Mode, take KEGmodel as input, and there is not much difference in input attributes. Both models contain knowledge graph entities (Num

Entity, Num Correlation, Hidden Dim, Gamma), Relationship hidden layer, etc. Therefore, in the HAKE model, two additional parameters are introduced for comparison with modulus. In addition, in terms of filter introduction rating and type selection after rating, weighting and phase related calculations are performed. The Func function is the key code location for hierarchical construction of the HAKE model. Through test analysis, it is found that the model calculates Mod by_ Relation, Bias_ The relationship calculation score control model generates entities at the same level. In order to achieve more entities of the same type at the same level, modify R_ Score, Phase_ Score calculation, based on HEAD_ BATCH header node, relationship with AIL_ The BATCH tail node, with a relationship similarity count as a parameter, is added to the calculation, and then filter parameters are generated.

The details are as follows:

a) Add Batch to the input parameters of the function body_ Sigle,
b) Calculate Dr. Dm distance parameter
c) Calculate Dr. Dp distance parameter
d) Import HEAD_ BATCH: (?, r, t), TAIL_ BATCH: (h, r, ?)
e) Batch_ Sigle calculates the similarity between head and tail entities and r
f) Calculate R_ Score, Phase_ Score
g) Generate Filter Filter Parameters
h) Entity attribute annotation, entity type comparison with Filter return value, weighted calculation
i) Data entity generation call

After the above steps, the probability of data generation for similar entities can be enhanced.

4   Image Classification and Fusion Mode Based on Knowledge Atlas

4.1   Comparison of Optimization Plans

Model optimization is often involved in image classification tasks, and there are several main aspects for optimizing convolutional neural networks, including accuracy optimization, memory optimization, and training time optimization. The main purpose of this paper is to improve the accuracy of the model. Therefore, the image fusion mode will be compared with other convolutional neural network optimization schemes to analyze the advantages of the fusion mode.

Common model optimization methods include optimizing data and optimizing the structure of convolutional neural networks. The optimization of data includes data cleaning and data enhancement, which essentially improve the quality of the dataset. Data cleaning can reduce erroneous and invalid data. Incorrect data can help data learn incorrect parameters and reduce classification accuracy. Increasing data can fully train model parameters, Especially in large models, where there are many layers of the network and the amount of data is too small, it can lead to underfitting during model training. Therefore, data augmentation techniques can be used to rotate, clip, invert, scale, shift, and

reduce Gaussian noise on data, such as image data. Even generate adversarial network GANs. For the problem of data as a whole, for example, if the data volume is small, you can use the oversampling method to repeatedly sample the insufficient categories, or SMOT can generate less category data based on similarity to increase the weight of less sample data. For a large amount of data, we can use undersampling to prevent model overfitting, cluster first, then randomly sample the large class, or randomly sample first, and then use Boosting algorithm to integrate weak classifiers to generate strong classifiers. However, there is a problem with optimizing the data aspect, which is through data optimization. There are always limitations to the model itself, and the degree of data optimization improvement is not infinite.

The network optimization of convolutional neural networks is relatively rich, and the development history of convolutional neural networks mainly revolves around the optimization of network structure. The optimization of network structure includes convolutional kernel optimization. Generally, the larger the convolutional kernel, the higher the accuracy, but it will consume memory. Moreover, due to the complex connection between features of convolutional kernels, the generalization performance of the model will deteriorate, with ResNet and VGGNet being representatives. Optimizing network size and increasing network width will enable each layer of the network to learn more features, and increasing depth will also improve accuracy, but the training process will become slower. Activation function optimization. ReLU, Sigmaid, ELU, PReLU, Sigmaid, or LeakyReLU activation function can be used to improve the effect. Optimizer optimization, including SGD, Adam, etc., requires repeated attempts.

Through analyzing data optimization and network structure optimization, we have found that any optimization to the extreme cannot exceed the limitations of the model itself. For example, processors, memory, and even faster processors have operational limitations, requiring memory coordination when the calculation data is too large. The development of models is parallel to the development of knowledge graphs. Returning to the knowledge graph and comparing image classification algorithms, knowledge graphs are like external storage that stores a large amount of resource information. Image classification algorithms cannot obtain information such as pain, lifestyle habits, past illnesses, etc. They need to be mutually verified in the knowledge graph. The results of image classification are scoped on a knowledge graph, which disassociates certain results of image classification and ultimately reduces the classification range to 1, resulting in correct results. Therefore, there is no doubt that the combination of the two will bring better results. As a whole, the key to the fusion mode is to put the part of the image classification results with low accuracy into the knowledge map for clustering analysis, and then calculate the overall probability, that is, the samples that were judged correctly by the convolutional neural network continue to remain, and the uncertain samples are handed over to the knowledge map for judgment, so as to improve the classification accuracy.

Usually, image classification tasks, especially multi classification tasks, use probability distributions to represent the calculation results of each category, while binary classification outputs as 01. Multi classification adds a SoftMax() function at the end of the network to obtain probability values. The commonly used indicators for binary evaluation include accuracy, recall, and F1. Multi classification evaluation indicators

include Micro F1 and Macro F1. In addition to binary classification and multi classification, there is also multi label classification, where each sample can be predicted as one or more categories.

From previous research, it can be seen that the solution to any problem is a gradual decomposition of the problem, breaking it into smaller parts. The solution to the multi classification problem is also the same. Multi classification can be converted into an OvO pairwise combination and converted into N (N-1)/2 binary classification tasks, and the results can be voted out; Train N classifiers with OvRn combinations, with the highest confidence in the selection; Multiple labels and multiple classifications can also be decomposed into N binary classification problems. However, if such decomposition is done directly, when the total category is greater than the number of labels to be selected, there will be category imbalance, as there are too many categories, leading to imbalance between positive and negative samples. So a second idea emerged, whether it is possible to transform the multi label problem into a multi classification problem and then into a binary classification problem. Based on the previous understanding of the problem of knowledge map construction, we are very familiar with the concept of mapping. The idea of fusion mode in this chapter is to reduce the dimension of multi label problem to multi classification problem, and then map the problem to the knowledge map for solution, and gradually solve it in the knowledge map.

Returning to the multi classification problem to see if there are specific solutions, the solution of the multi classification problem

In image classification, there are two types of tasks: multi classification tasks and binary classification tasks. However, overall, multi classification tasks are decomposed into binary classification tasks for processing. The decomposition from multi classification tasks to binary classification tasks involves dividing them into scatter points for voting, direct N classification with the highest confidence, and constructing a binary classification classifier with a graph to calculate distance. Another method is based on hierarchical classification, which calculates confidence through tree structure.

Overall, this article believes that although image classification, especially disease classification, usually adopts multi classification methods, in fact, disease image classification should be considered as multi label classification from the perspective of images, because some diseases and diseases cover the symptoms of another disease, which are similar or subclasses themselves. Therefore, multi classification, a single-layer classification result, is not very suitable for disease classification diagnosis, Therefore, introducing hierarchical attributes into the classification results can better determine the level of the disease and thus determine its name.

4.2   Gambling Tree

Based on the concept of hierarchical two classifiers mentioned and the domain limited knowledge map built based on the improved hierarchical awareness knowledge map in this paper, we can naturally think about whether we can use the knowledge map with hierarchical characteristics to optimize classification and generate computing strategies based on the classification results of hierarchical two classifiers. Because nodes are tree structure and hierarchical structure, computing strategies are generated by computing trees [13–15].

Therefore, this article proposes the concept of game tree, including credit parameter Cre (Credict), power parameter Mig (Migt), and inference parameter Ill (Illation). Cre represents the degree of trustworthiness of the preliminary results of image classification within domain limitations. The Silhouette Coefficent Si and the probability Pi $\{1,2,3…i\}$ of entities in multiple classification domains are used to calculate the aggregation degree formula based on the entities in the domain.

The formula for Cre is expressed as:

$$Cre(i) = \frac{O(i) - I(i)}{max(I(i), O(i))}$$

$$Cre(i) = \begin{cases} 1 - \dfrac{I(i)}{O(i)} & I(i) < O(i) \\ 0 & I(i) = O(i) \\ \dfrac{O(i)}{I(i)} - 1 & I(i) > O(i) \end{cases}$$

In addition, the formulas for I (i) and O (i) are expressed as:

$$I(i) = \frac{1}{n-1} \sum_{j \neq i}^{n} distance(i,j)/PiPj$$

$$O(m) = \frac{1}{n-1} \sum_{j \neq i}^{n} distance(i,j)/PiPj$$

$$O(i) = Min\{O_1(i)O_2(i)O_3(i)...O_m(i)\}$$

I represents other sample points within the same class as sample j, and Distance represents the distance between i and j. So the more novel I (i) is, the closer it becomes.

The calculation method of O (i) is similar to that of I (i). Just need to traverse other class clusters to obtain multiple values and select the smallest one from them.

The strength parameter Mig represents the degree of possible difference between the classification result with the highest probability among multiple classification results and other classification results. When the strength parameter is greater than the threshold, domain restrictions can be ignored. The setting concept of referring to the real world is that regardless of the difference in symptoms described by patients, the results of physical and chemical tests still prevail.

$$Mig(i) = \frac{1}{n-1} \sum_{j \neq i}^{n} ((Pi - Pj)$$

The inference parameter Ill correctly represents the score obtained by combining knowledge inference with multiple classification probabilities when the strong parameters outside the domain fail and the strong parameters within the domain also fail. In cases where the probability of multi classification is generally low for inference parameters, more reliable data information comes from the knowledge graph, which selects the optimal classification result based on the number of knowledge inference question answering entities and edge links.

$$Ill(i) = Max\{(M(i)+N(i)), M(i+1)+N(i+1)),...M(n)+N(n)\}$$

Calculate Path

Game trees can generate image classification models that utilize information from knowledge graphs. The input stack of the game tree is the classifier probability of all classification results, as well as the entity nodes and relationships represented in the domain restricted knowledge graph. Firstly, the strength parameters of all classification results are calculated from the root node, and if they are greater than the threshold, the results are output. Otherwise, match the domain restricted knowledge graph. If the maximum probability is not within the domain, delete the maximum classification result in the stack, and then calculate the domain strong parameter. If it is greater than the threshold, output the result. Otherwise, calculate the maximum possible result in the domain through knowledge reasoning. Calculate the game value when the input is a multi graph input:

Gambling = Mig\ast Ill + Cre\funcapply(i)-\beta

Output the results with large values.

KG based CNN image classification fusion model based on domain limited knowledge map

There are several task scenarios for the fusion of knowledge graph and image classification:

① Only image classification data is input. When only image classification data is input in the fusion model, it is impossible to generate a domain limited knowledge map, but there is still a hierarchical perception knowledge map that can be used to calculate the map. The principle of the fusion mode is to project the uncertain classification of image recognition into the map for calculation. When the map range is large and not accurate enough, it can control to reduce the threshold value of the powerful parameter Mig, Enable probability selection conditions to take effect through threshold control.

② Image input and low confidence/high confidence Q&A knowledge input, when there is image input and low confidence Q&A knowledge input, it means that there will be domain restrictions on knowledge graph generation with low confidence. However, due to probability threshold control, the overall classification accuracy will not change much, as the graph only affects data with uncertain image classification algorithms. When high confidence Q&A knowledge is input, the aggregation of entities projected into the knowledge graph will be very high, and the overall game value will change to make it the output classification result. In theory, when the knowledge graph is detailed enough and the threshold is lowered, the classification result will approach complete accuracy.

③ Multi graph input and question answering knowledge input. When there are multiple graph data inputs, the real multi graph data may belong to different categories or the same category. If all the multiple graphs are input into the model, multiple results will be calculated in the game tree, and the strength parameters will be calculated on the results. If it is greater than the threshold, one result will be output, and if it is less than the threshold, multiple results will be output.

In disease image classification tasks, the usual decision tree input is the corresponding disease entity, but symptom entity parts such as hands, feet, red, purple, rules, blur, and other information can also be inputted [16–18]. Target detection can be used for recognition, and sometimes there may be no input from the question answering system.

Therefore, this paper expands the game tree to an image classification fusion mode KG based CNN that is more suitable for various situations to deal with various use scenarios.

Image input can use object detection algorithms for entity extraction, followed by knowledge graph comparison, game tree calculation, and finally output the best classification result [19]. The specific fusion mode has the following levels: the first is the input layer, which is image data and knowledge question and answer data (knowledge question and answer data can be empty). The question and answer data is entities and entity attributes filled with a value of 0/1. Connected with the input layer is the convolutional neural network model layer. The model layer pre trains the classification data to output the classification probability value softmaxt, and then the knowledge map layer, Project the entities onto the knowledge map (when there is no question and answer data input, cluster the entities and calculate the entity distance) (when there is question and answer data input, knowledge extraction is conducted according to the input entities to form a domain limited knowledge map, and the entities and probability obtained by classification are projected onto this map), carry out the calculation path of the game tree, when multiple maps are input, carry out the calculation of multiple game trees, and finally output the classification results.

5   Experiment and Analysis

5.1   Evaluating Indicator

The evaluation index used in this paper is Accuracy, true positive (TP): melanoma is diagnosed, but actually there is also melanoma. False positive (FP): diagnosed as melanoma, but not actually. True negative (TN): there is no melanoma diagnosed, and there is no melanoma in fact. False negative (FN): there is no melanoma, but there is. Due to the unreliable accuracy evaluation when data is abnormally imbalanced, several other indicators are introduced for comprehensive evaluation. Precision focuses on evaluating how much of the data predicted to be positive is actually positive? Recall focuses on evaluating how much of all Positive data has been successfully predicted as Positive

Calculation formula:

$Accuracy = TP + FN/TP + FP + TN + FN$

$Precision = TP/TP + FP$

$Recall = FN / TN + FN$

$F1\ score = 2\ Precision* Recall / Precision + Recall$

The confusion matrix is usually used as the evaluation standard of model performance in multi classification, and the confusion matrix can describe the gap between the predicted value and the real value. Specifically, it is to calculate the prediction of each category in a certain category. It is represented by a matrix representation. The horizontal and vertical coordinates are multi category categories. Each horizontal row represents the prediction value of each category in the corresponding vertical coordinate category. Observing the performance of the model in various categories can calculate the accuracy and recall of the model corresponding to each category. Through the confusion matrix, we can observe which categories are not easy to distinguish directly, such as how many of category A are classified into category B, which can be targeted to design

features, making categories more differentiated. Simply put, it's about seeing how many misjudgments there are.

HAM10000 has 10000 image data, with 6000 as model training data and 1500 as test data. From the remaining 2500, 1400 samples of seven balanced classifications (with the same number of samples for each classification) were extracted as validation data. As shown in Fig. 5.2, a confidence level of about 95% for sample size calculation is more appropriate (Fig. 8).

Fig. 8. Skin dataset HAM10000.

5.2   Performance Comparison Experiment of Fusion Pattern Classification Model

The experimental control model is ResNet, and the accuracy rate of the model is 68%. According to the reasoning in the previous summary, overfitting phenomenon should have occurred. To ensure the balance of samples, a lot of training set data has been sampled, generally 7000 training data. In addition, the training model in this paper does not use the pre training model for transfer learning, which may also lead to the reduction of classification accuracy, but the fusion model can still improve the accuracy of the low performance model. Compared to other situations, as shown in Fig. 5.5. It can be seen that the classification calculation strategy of fusion mode is applied without input, NAN_ Compared with ResNet, the accuracy rate of INPUT has increased to 104%. When the text data with confidence is input, the accuracy rate of INPUT is slightly improved compared with the method of directly applying knowledge map clustering without input. However, when the confidence is increased, the accuracy rate of single map input is improved to 108%. In addition, when conducting multi graph input data experiments, the accuracy significantly increased. Specifically, the multi graph experiment conditions were to input a total of 140 sets of test sample graphs, 420 images, and obtain a single graph conclusion through game tree analysis. The results were compared with other graph classification results for election voting. The minority followed the majority and then compared based on the Gambling value (Fig. 9).

Based on the decrease in accuracy of the experimental model mentioned earlier, this section obtained the model ResNet++ by retraining the model parameters. As shown in
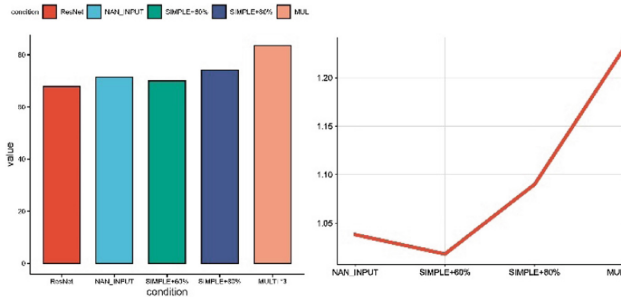
Fig. 9. Comparison of High and Low Confidence Multiple Input Performance

the figure, it can be seen from the data that the high-performance classification model has a higher accuracy improvement after applying the knowledge graph. In the SIMPLE and 80% confidence knowledge Q&A text input scenario, compared to the ResNet model, it has improved to 113%, with a classification accuracy of 77.21%. The confusion matrix of ResNet++ model is shown on the right side of the figure above. It can be seen that the accuracy of the model in BCC, NV and VASC is high, while other items are still scattered, but there is a general trend of bias towards a certain category. At the same time, it can be seen that the high probability part still maintains the same distribution as before using the fusion mode. Analyzing the reasons, the probability distribution of failure data in high-performance classification model testing may be difficult to disperse, and correction may be possible after game tree calculation. The low performance classification model is easy to keep the wrong classification unchanged when calculating the game tree due to the comparative calculation of overfitting probability distribution (Fig. 10).
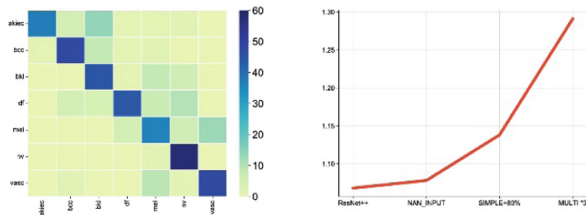

Fig. 10. High performance model ResNet + + confusion matrix and ratio curve with ResNet

5.3   Ablation Experiment

As shown in the figure, the accuracy comparison between ResNet and improved Res'net++ is shown. By analyzing the accuracy changes of ResNet and optimized ResNet++ without domain restricted knowledge graph conditions, and comparing the knowledge graph based on ordinary data extraction methods, it can be seen that the accuracy of domain restricted knowledge graph in the figure has been improved. Analyze the reasons, The main method for extracting non domain restricted knowledge graphs is to

iteratively extract entities from specified entity nodes, which can result in excessively dispersed knowledge and connections between various classified entities projected onto the knowledge graph. As a result, the final classification result with the highest weight is still the original classification result (Fig. 11).
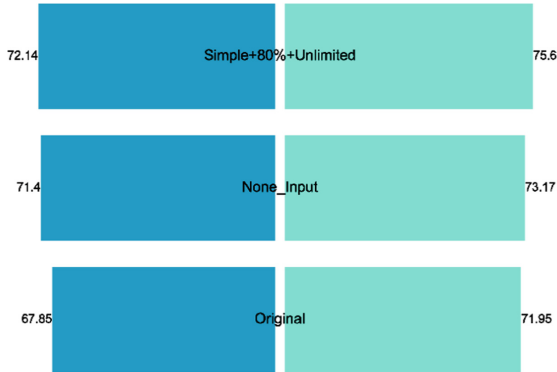

Fig. 11. ResNet and ResNet + + Domain Restricted Knowledge Graph Ablation Experiment SAMPLE ANALYSIS

As shown in the figure, the horizontal axis represents the classification of seven skin diseases, and the vertical axis represents the sample label, which is the case of single image input. The above figure shows a heat map, of which thirty samples were extracted from the HAM10000 dataset, and all the intercepted samples belong to Akiec. By observing the color depth, it can be determined that the probability values obtained by multiple classifiers Softmax in the image are still distributed among other categories. It is obvious that sample 19 is in a state of classification error, which is relatively evenly distributed among four categories. Generally, data with a probability distribution similar to this can be repositioned to the correct category by the knowledge graph in multiple sample data, but samples like sample 23 cannot be repositioned because the threshold of the strength parameter Mig is generally set to 0.4, and the difference between the maximum probability and other probabilities is 40%, This way, the calculation results will be returned at the first level of the game tree. However, setting the threshold too high may lead to classification errors in sample 13, as the distance between mel and Nv on the graph is very close, and the corresponding trusted values in the graph are also relatively large. Comparing the Gambling parameter will exclude the Akiec class from the computational space.

In general, the fusion mode can relocate some wrong samples to the correct position, and will only classify the original correctly classified samples into the wrong category with a very small probability. For example, the probability gap between the sample needs and other samples is less than the threshold, and then the probability sum of other classifications is greater than it, and the gap is further increased after the calculation of aggregation, execution, and reasoning values, This is a small probability because the calculation of these values is biased towards the correct classification (Fig. 12).
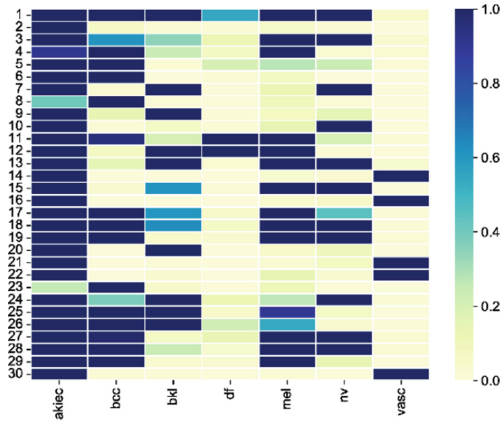
Fig. 12. Sample Thermogram

6   Conclusion and Future Work

In future research work, further exploration can be conducted on the construction of knowledge maps and diagnostic reasoning for skin diseases. Firstly, the application of knowledge graphs can help doctors better understand the causes, symptoms, treatment methods, and other aspects of skin diseases. We can consider introducing more multimodal data sources, such as videos, text, etc., to expand the breadth and depth of the knowledge graph and explore new medical knowledge from it. For example, by structuring, linking and representing the existing literature, cases, drugs, pathological images and other data, a more complete and systematic knowledge system of skin diseases can be built. You can try to deeply explore the correlation and regularity between medical data based on more complex and complete data, combined with medical images and laboratory test results. This knowledge map can be applied to natural language processing, image analysis, intelligent question and answer and other aspects to help doctors diagnose and treat more quickly and accurately.

References

1. Ridell, P., Spett, H.: Training Set Size for Skin Cancer Classification Using Google's Inception v3 (2017)
2. Alabduljabbar, R., Alshamlan, H.: Intelligent multiclass skin cancer detection using convolution neural networks. (010):000 (2021)
3. Meng, T., Lin, L., Shyu, M.L., Chen, S.C.: Histology image classification using supervised classification and multimodal fusion. In: 2010 IEEE International Symposium on Multimedia, Taichung, Taiwan, pp. 145–152 (2010).https://doi.org/10.1109/ISM.2010.29
4. Znaidia, A., Shabou, A., Popescu, A., et al.: Multimodal feature generation framework for semantic image classification. In: ACM International Conference on Multimedia Retrieval, pp. 1–8. ACM (2012)
5. Ji, S., Pan, S., Cambria, E., et al.: A survey on knowledge graphs: representation, acquisition, and applications. IEEE Trans. Neural Netw. Learn. Syst. (99) (2021)

6. Zhen, W., Zhang, J., Feng, J., et al.: Knowledge graph embedding by translating on hyperplanes. In: National Conference on Artificial Intelligence. AAAI Press (2014)
7. Feng, J.: Knowledge graph embedding by translating on hyperplanes. In: AAAI, vol. 28, no. 1 (2014)
8. Moon, C., Harenberg, S., Slankas, J., et al.: Learning contextual embeddings for knowledge graph completion. In: The 21st Pacific Asia Conference on Information Systems, vol. 10 (2017)
9. Ji, G., He, S., Xu, L., et al.: Knowledge graph embedding via dynamic mapping matrix. In: Meeting of the Association for Computational Linguistics & the International Joint Conference on Natural Language Processing, pp. 687–696 (2015)
10. Dettmers, T., Minervini, P., Stenetorp, P., et al.: Convolutional 2D Knowledge Graph Embeddings (2017)
11. Nguyen, D.Q., Vu, T., Nguyen, T.D., et al.: A capsule network-based embedding model for knowledge graph completion and search personalization (2018)
12. Yao, L., Mao, C., Luo, Y.: KG-BERT: BERT for knowledge graph completion (2019)
13. Nordhausen, K.: An introduction to statistical learning—with applications in R by Gareth James, Daniela Witten, Trevor Hastie & Robert Tibshirani. Int. Stat. Rev. **82**(1), 156–157 (2014)
14. Wang, H., Zhang, F., Xie, X., et al.: DKN: deep knowledgeaware network for news recommendation. In: Proceedings of the 2018 World Wide Web Conference on World Wide Web, pp. 1835–1844 (2018)
15. Wang, Q., Mao, Z., Wang, B., et al.: Knowledge graph embedding: a survey of approaches and applications. IEEE Trans. Knowl. Data Eng. **29**(12), 2724–2743 (2017)
16. Liu, Y., Li, H., GarciaDuran, A., et al.: MMKG: multimodal knowledge graphs. In: European Se mantic Web Conference, pp. 459–474 (2019)
17. MoussellySergieh, H., Botschen, T., Gurevych, I., et al.: A multimodal translationbased approach for knowledge graph representation learning. In: Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, New Orleans, Louisiana, pp. 225–234 (2018)
18. Cun, Y.L., Boser, B., Denker, J.S., et al.: Handwritten digit recognition with a backpropagation network. Adv. Neural. Inf. Process. Syst. **2**(2), 396–404 (1990)
19. Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. **25**(2) (2012)