INT. J. BIOAUTOMATION, 2021, **25**(1), 87-100 doi: 10.7546/ijba.2021.25.1.000812

Disease Diagnosis of Dairy Cow by Deep Learning Based on Knowledge Graph and Transfer Learning

Meng Gao[1], Haodong Wang[1], Weizheng Shen[1], Zhongbin Su[1,*], Huihuan Liu[2], Yanling Yin[1], Yonggen Zhang[3], Yi Zhang[4]

[1]*College of Electrical Engineering and Information*
Northeast Agricultural University
Harbin 150030, China
E-mails: gaomeng@neau.edu.cn, whaodong1997@163.com,
wzshen@neau.edu.cn, suzb001@163.com, yinyanling@neau.edu.cn

[2]*Dairy Association of Heilongjiang Province*
Harbin 150020, China
E-mail: liuhuihuan@sina.com

[3]*College of Animal Science and Technology*
Northeast Agricultural University
Harbin 150030, China
E-mail: zhangyonggen@sina.com

[4]*College of Computer Science and Electrical Engineering*
East University of Heilongjiang
Harbin 150066, China
E-mail: chrisrang@163.com

[*]*Corresponding author*

Received: June 08, 2020 **Accepted: February 19, 2021**

Published: March 31, 2021

Abstract: *In dairy herd management, it is significant and irreplaceable for veterinarians to make rapid and effective diagnosis of dairy cow diseases. Based on electronic medical records, deep learning (DL) has been widely used to support clinical decisions for humans. However, this method is rarely adopted in veterinary diagnosis. In addition, most DL models are driven by large datasets, failing to utilize the knowledge acquired by veterinarians in subjective experience, which is critical to disease diagnosis. To address these problems, this paper proposes a DL method for disease diagnosis of dairy cow: convolutional neural network (CNN) based on knowledge graph and transfer learning (KGTL_CNN). Firstly, the structural knowledge was extracted from a knowledge graph of dairy cow diseases, and treated as part of the inputs to the CNN based on knowledge graph (KG_CNN). Then, the model performance was enhanced through pre-training by transfer learning. To verify its performance, experiments were carried out on dairy cow clinical datasets. The results show that our model performed satisfactorily on disease diagnosis: the KG_CNN and KGTL_CNN achieved an F1-score of 85.87% and 86.77%, respectively, higher than that of typical CNN by 6.58% and 7.7%. The research results greatly promote the effective, fast, and automatic clinical diagnosis of dairy cow diseases.*

Keywords: *Knowledge graph, Deep learning, Transfer learning, Dairy cow, Disease diagnosis.*

Introduction

In livestock farming, it is of great significance to prewar and diagnose dairy cow diseases rapidly and effectively. The early warning and diagnosis help to implement targeted treatment as soon as possible, thereby improving the health of dairy herds. In recent years, many wearable and non-contact devices have been used to monitor and analyze individual behaviors of dairy cows, e.g., estrus [7, 20, 22], lameness [16, 27], rumination [9, 11, 28], and feeding [9], laying solid theoretical basis for early warning of diseases.

Early warning is for timely diagnosis and treatment, and an effective diagnosis usually requires professional veterinarians to make judgments based on clinical symptoms and relevant examinations, whose importance is irreplaceable. Disease diagnosis for dairy cow is somewhat similar with that for human. But the primary diagnosis is rather complex, because animals cannot directly tell their feelings. With more knowledge and experience, general practitioners are more suitable for diagnosing diary cow diseases than specialists. However, due to the lack of knowledge, even an experienced veterinarian might misdiagnose uncomplicated diseases.

Expert system is commonly used to assist with the diagnosis of diseases [2, 5, 6, 10, 14, 29]. Gao et al. [8] developed an ontology-based knowledge representation model for diagnosing equine diseases. Nusai et al. [25] established an uncertain knowledge representation model for swine disease diagnosis. Hamedan et al. [12] created set of fuzzy rules for predicting chronic kidney disease. Nugroho [24] applied the fuzzy Tsukamoto method to evaluate the risk level of endometritis disease in cattle. Because of their reasoning mechanism, the above methods are intuitive and explainable. However, they are complicated by the large number of high-quality and necessary rules or cases, and difficult to mine the hidden relations among clinical symptoms.

At present, machine learning (ML) has performed well in diagnosing human diseases. The disease diagnosis model based on clinic medical records is increasingly meaningful in primary diagnosis. Maini et al. [23] applied five ML algorithms, including k-nearest neighbors (KNN), Naïve Bayes (NB), logistic regression (LR), adaptive boosting (AdaBoost), and random forest (RF), to predict early cardiovascular diseases based on the medical records of South India. Zhao et al. [31] relied on an RF model to forecast chronic kidney diseases. These methods are proved to be efficient, yet their performance hinges on the manually extracted features.

Recently, deep learning (DL) has gained popularity in disease diagnosis. Ljubic et al. [19] combined long short-term memory (LSTM) and recurrent neural network (RNN) into a DL model to judge whether a patient will get Alzheimer's disease. Pham et al. [26] built an LSTM model to predict healthcare trajectories from medical records. Due to the automatic feature extraction from massive labeled data, these DL methods are effective in disease diagnosis. Unfortunately, due to the lack of diagnosis experience of experts, the DL methods cannot effectively find the hidden correlations between diseases and symptoms that are not included in medical records.

To address the problem, this paper proposes a DL-based intelligent disease diagnosis method for dairy cow, driven by veterinary knowledge and medical records. Specifically, the experimental knowledge was extracted from a knowledge graph oriented to dairy cow diseases, and used to guide the learning process of convolutional neural network (CNN). Next, our model was pretrained through transfer learning to achieve better performance on a limited number of real-world samples.

The remain of this paper is organized as follows: Section 2 describes our datasets and the proposed method; Section 3 carries out comprehensive experiments, and analyzes the experimental results; Section 4 summarizes the research findings.

Methodology

As shown in Fig. 1, our method primarily consists of three parts: data processing, CNN based on knowledge graph (KG_CNN), and CNN based on knowledge graph and transfer learning (KGTL_CNN).

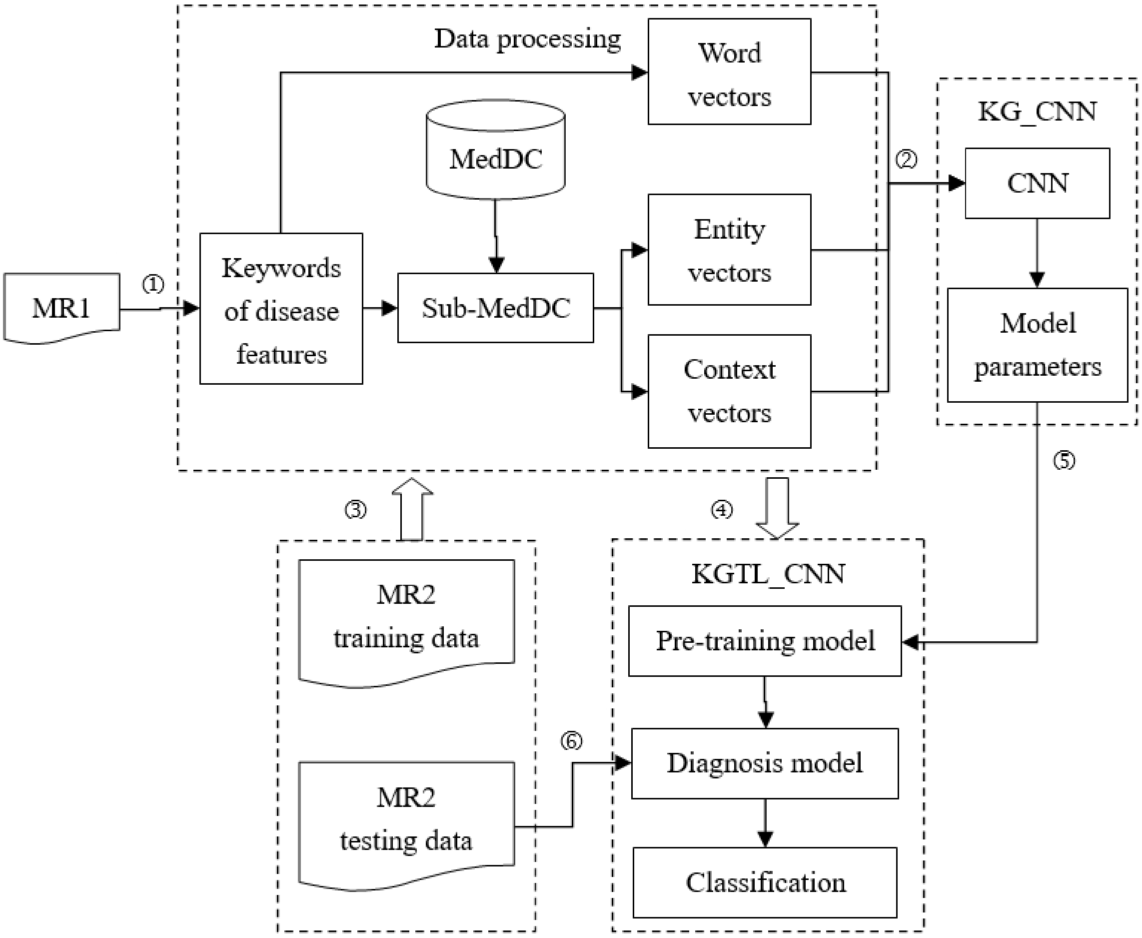

Fig. 1 Flow chart of our method, where MR1 is the generated dataset;
MR2 is the real-world dataset;
MedDC is the medical knowledge graph oriented to dairy cow;
Sub-MedDC is the subgraph of MedDC.

Firstly, the MR1 dataset was processed by the data processing module, and the vectors representing features and relevant knowledge were obtained. Secondly, the KG_CNN model was pretrained on the processed dataset, and the model parameters were obtained. Then, the MR2 dataset was divided into training data and testing data, and they were processed by the data processing module, respectively. Next, the KGTL_CNN model was trained on the training data of the MR2 dataset using the pre-trained model, and the diagnosis model was obtained after fine tuning parameters. Finally, the diagnosis model was validated by the testing data of the MR2 dataset.

Apart from the features of medical records, our model also utilizes the experimental knowledge between symptoms and diseases of dairy cow. In addition, our model can effectively learn from a limited number of samples through transfer learning.

Datasets

Two medical record sets of dairy cow, namely MR1 and MR2, were adopted for this work. MR1 was used to pretrain the KG_CNN, and MR2 to train and test KGTL_CNN. Table 1 presents further details and compares the two datasets.

Table 1. Description of the datasets

Number of medical records		**MR1**	**MR2**
		18,580	3,069
Types of diseases	Mastitis	2,816 (15.16%)	920 (29.98%)
	Forestomach atony	2,679 (14.42%)	482 (15.71%)
	Rumen indigestion	3,084 (16.60%)	373 (12.15%)
	Gastroenteritis	3,109 (16.73%)	762 (24.83%)
	Urmen acidosis	3,252 (17.50%)	281 (9.16%)
	Abomasum dislocation	3,640 (19.59%)	251 (8.18%)

(1) MR1 Datasets

The MR1 dataset was manually generated in the following steps [30]: first, the descriptions of various dairy cow diseases, including symptoms, pathological changes, and differential diagnoses, were extracted from professional books, and compiled into a standard medical record for each disease. According to punctuations and degree of association between contexts, each standard medical record was split into several short sentences. For each set of sentences, half of them were selected each time by method of simple random sampling without replacement [30], and organized into a medical record of dairy cow. Through multiple sampling, a large number of visual medical records were collected for different diseases.

(2) MR2 Datasets

The MR2 dataset includes real-world medical records of 832 dairy cows from 13 large-scaled breeding farms in Heilongjiang, China. These records have been maintained by Dairy Association of Heilongjiang Province (DAHLJ), one of our partners. Each medical record contains clinical symptoms, laboratory test, diagnosis, disposal measure, medication use, and outcome after treatment.

Data processing

Prior to applying the pre-training model, the MR1 dataset was preprocessed, and the MR2 dataset would also be preprocessed before used to train the proposed model. As shown in Fig. 1, the data processing includes feature selection and knowledge extraction.

(1) Feature selection

The medical record of each dairy cow was processed on the Language Technology Platform (LTP) of Harbin Institute of Technology (HIT) [4], using the professional dictionary of diseases and symptoms. Every medical record went through the extraction of feature keywords, text segmentation, part-of-speech (POS) tagging, recognition of named entities, and the removal of stop words and meaningless words. In this way, a set of keywords about disease features was obtained as $x = [v_1, v_2, \dots, v_n]$, where n is the number of keywords and v_i is the i-th keyword.

Then, all the extracted features were converted into a vector matrix. The skip-gram model of Word2Vec [13] was adopted to transform each v_i into its vector representation by looking up the embedding matrix $\boldsymbol{v_i} \in \mathbb{R}^k$, where k is the dimensions of the word vector.

Let $\boldsymbol{X}_{\boldsymbol{1}:\boldsymbol{n}} \in \mathbb{R}^{k\times D} = \boldsymbol{v}_{\boldsymbol{1}} \oplus \boldsymbol{v}_{\boldsymbol{2}} \oplus \ldots \oplus \boldsymbol{v}_{\boldsymbol{n}}$ be the embedding matrix of a medical record, where D is the number of features.

(2) Knowledge extraction

Based on the keywords acquired in the previous subsection, the relevant knowledge was extracted from disease knowledge graph of dairy cow, and transformed into vector representations, too.

As shown in Fig. 2, a knowledge graph MedDC typically contains empirical facts and relationships between symptoms and diseases of dairy cow. It can be denoted as a large set of triples (entity(head)-relation-entity(tail)): MedDC $= \langle h, r, t\rangle$, where $h \in \mathcal{E}$, $r \in \mathcal{R}$, and $t \in \mathcal{E}$ are the head, relation, and tail of a knowledge triple, respectively. The set of entities and set of relations in the MedDC are denoted as $\mathcal{E}$ and $\mathcal{R}$, respectively.

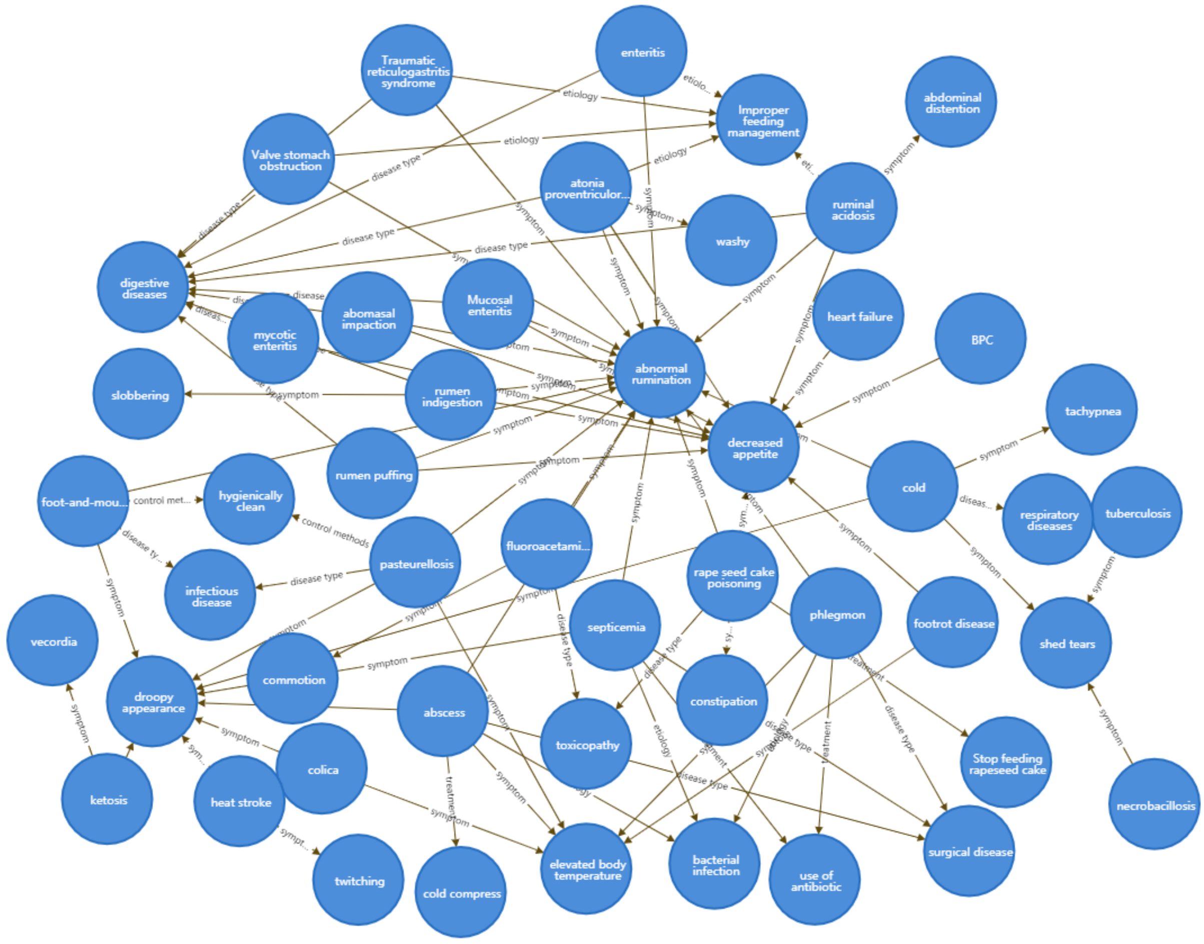

Fig. 2 Illustration of the knowledge graph MedDC

Given the embedding matrix $\boldsymbol{X}_{\boldsymbol{1}:\boldsymbol{n}}$ of a medical record, the similarity between $\boldsymbol{v}_{\boldsymbol{i}}(i = 1,2, \ldots, n)$ and entity h or t in MedDC can be calculated by entity linking [21]. The set of i relevant entities can be defined as:

$$\varepsilon = \{e_i | e_i = argmax_{e\in MedDC} Score(e, \boldsymbol{v}_{\boldsymbol{i}}), i = 1,2, \ldots, n\},$$

where $Score(e, \boldsymbol{v}_{\boldsymbol{i}})$ is the cosine similarity between v_i and an entity e in MedDC.

For the input entity $e_i \in \varepsilon$, the 1-hop ripple set can be extracted to form a sub knowledge graph:

$$S_i = \{(h, r, t) | (h, r, t) \in MedDC \text{ and } (h \in e_i \; or \; t \in e_i)\}, i = 1,2, \dots, n.$$

Suppose $\varepsilon(e_i) = \{e | e \in (h, t) \text{ and } ((h, r, e_i) \in S_i \; or \; (e_i, r, t) \in S_i)\}$ is the set of context entities of e_i, with m being the number of entities of $\varepsilon(e_i)$; $e_{ij} \in \varepsilon(e_i)$ ($j = 1,2, \dots, m$) is the j-th relevant entity of e_i; $\boldsymbol{e_{ij}}$ is the embedding vector of e_{ij}, which was obtained by knowledge representation learning using the TransD model [15].

The semantic similarity between words is positively correlated with the similarity between the real number vectors mapped in the same space vector. Therefore, the context vector can be represented by the mean of all context entities of e_i:

$$\overline{\boldsymbol{e_i}} = \frac{1}{m} \sum_{e_{ij} \in \varepsilon(e_i)} \boldsymbol{e_{ij}}, i = 1,2, \dots, n, j = 1,2, \dots, m.$$

KG_CNN

For each feature v_i, its entity vector can be expressed as $\boldsymbol{e_i} \in \mathbb{R}^{d \times 1}$, and the context vector as $\overline{\boldsymbol{e_i}} \in \mathbb{R}^{d \times 1}$,where d is the dimensions of entity vector. Let $q(\boldsymbol{e_i})$ be the vector of $\boldsymbol{e_i}$ after entity alignment, and $\boldsymbol{Q_{1:n}}$ be the aligned entity vector of a medical record:

$$\boldsymbol{Q_{1:n}} = q(\boldsymbol{e_1}) \oplus q(\boldsymbol{e_2}) \oplus \dots \oplus q(\boldsymbol{e_n}).$$

Let $q(\overline{\boldsymbol{e_i}})$ be the context vector of $\overline{\boldsymbol{e_i}}$ after entity alignment, and $\overline{\boldsymbol{Q_{1:n}}}$ be the aligned context vector of a medical record:

$$\overline{\boldsymbol{Q_{1:n}}} = q(\overline{\boldsymbol{e_1}}) \oplus q(\overline{\boldsymbol{e_2}}) \oplus \dots \oplus q(\overline{\boldsymbol{e_n}}),$$

where q is the transfer function for word-entity alignment:

$$q(e) = tan(\boldsymbol{M}e + b),$$

where $\boldsymbol{M} \in \mathbb{R}^{k \times d}$ is the transfer matrix; $b \in \mathbb{R}^{k \times 1}$ is the deviation.

Herein, $\boldsymbol{X_{1:n}}$, $\boldsymbol{Q_{1:n}}$, and $\overline{\boldsymbol{Q_{1:n}}}$ are of the same dimension. The three vector matrices were aligned and superposed, and then imported to our CNN model as the multi-channel inputs:

$$\boldsymbol{X} = [\boldsymbol{v_1} q(\boldsymbol{e_1}) q(\overline{\boldsymbol{e_1}})][\boldsymbol{v_2} q(\boldsymbol{e_2}) q(\overline{\boldsymbol{e_2}})] \dots [\boldsymbol{v_n} q(\boldsymbol{e_n}) q(\overline{\boldsymbol{e_n}})] \in \mathbb{R}^{k \times n \times 3}.$$

To extract sufficient features of $\boldsymbol{X}$, several convolution kernels with different window sizes were employed to acquire more semantic information. Let $\boldsymbol{w} \in \mathbb{R}^{h \times h}$ be a convolution kernel, where h is the size of $\boldsymbol{w}$, and $\boldsymbol{X_{i:i+p-1}}$ be the vector matrix from v_i to v_{i+p-1}. Then, the result of $\boldsymbol{w}$ acting on $\boldsymbol{X_{i:i+p-1}}$ can be defined as:

$$C_i^t = f(\boldsymbol{w}\boldsymbol{X_{i:i+p-1}} + b),$$

where $f(*)$ is the rectified linear unit (ReLU) function [13]; b is the deviation; t is the serial number of convolution kernels.

Then, max-over-time pooling [32] was performed to select the maximum of the output feature map:

$$\tilde{C}^t = max\{C_1^t, C_2^t, \dots, C_{n-p+1}^t\}.$$

Combining the responses of all the $\tilde{C}^t$, the representation of a medical record x can be obtained as:

$$\boldsymbol{e}(\boldsymbol{x}) = \tilde{C}^1 \oplus \tilde{C}^2 \oplus \dots \oplus \tilde{C}^l,$$

where l is the number of convolution kernels.

Let p_r be the relevance probability by comparing $\boldsymbol{e}(\boldsymbol{x})$ to the r-th disease. After being normalized by Softmax, the p_r can be expressed as:

$$p_r = Softmax(s_r e_r(x) + b_r) = \frac{exp(s_r \cdot e_r(x) + b_r)}{\sum_{i=1}^{n} exp(s_i \cdot e_i(x) + b_i)},$$

where s_i and b_i are the parameter and deviation corresponding to the i-th disease, respectively; n is the number of disease types.

Here, adaptive moment estimation (ADAM) [1] is adopted to minimize the objective function in model training. The network parameters were updated iteratively through error back propagation until the model met the fitting requirements.

KGTL_CNN

As shown in Table 1, the MR2 dataset is much smaller than the MR1 dataset. Thus, this dataset cannot be directly applied to train the DL model. To solve the problem, the performance of our disease diagnosis model was improved through transfer learning [18].

Firstly, the MR2 dataset was processed by feature selection and knowledge extraction, producing the word vector, entity vector and context vector of each real-world medical records of dairy cow. Secondly, 80% of the dataset were randomly selected, and imported to the pre-trained KG_CNN, while the remaining 20% were treated as the test data. Thirdly, the network parameters were fine-tuned, and the softmax layer of KG_CNN was replaced with a fully connected layer to re-train the network. Finally, the KGTL_CNN was obtained after constant iterations and updates.

Performance measures

The performance of our method was evaluated in terms of Accuracy, Precision, Recall, and F1:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \times 100\%,$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \times 100\%,$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \times 100\%,$$

$$\mathrm{F1} = \frac{2\times\mathrm{Precison}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}} \times 100\%,$$

where TP and TN are the numbers of positive and negative samples that are correctly classified, respectively; FP is the number of negative samples that are incorrectly classified into positive samples; FN is the number of positive samples that are incorrectly classified into negative samples.

Experiments and results

Several experiments were conducted to verify the importance of each component of our model. The models without transfer learning were evaluated on real-world dataset, whereas the other models were evaluated on both the generated dataset and real-world dataset. The results of our model were compared with those of state-of-the-art models. The hyperparameters of our model were tuned through cross validation.

Our model

Our model was compared with the state-of-the-art models that have been applied in disease diagnosis. The contrastive models were divided into two groups: traditional models, and neural models. The performance comparison is shown in Table 2.

Table 2. Performance of our model and the state-of-the-art models

	Models	**Accuracy, (%)**	**Precision, (%)**	**Recall, (%)**	**F1, (%)**
Traditional model	SVM	75.13	72.46	74.83	73.63
	RF	72.83	72.19	71.63	71.91
	DT	71.53	71.70	68.32	69.97
Neural model	RNN	80.65	80.56	79.25	79.90
	CNN	81.22	80.93	80.21	80.57
Our model	KG_CNN	**86.22**	**85.52**	**86.22**	**85.87**
	TL_CNN	**84.55**	**84.74**	**83.14**	**83.93**
	KGTL_CNN	**87.04**	**86.02**	**87.53**	**86.77**

Note: SVM – support vector machine; RF – random forest; DT – decision tree; RNN – recurrent neural network; CNN – convolutional neural network; KG_CNN – CNN-based on knowledge graph; TL_CNN – CNN-based on transfer learning; KGTL_CNN – CNN-based on knowledge graph and transfer learning.

As shown in Table 2, the traditional models did not perform well, due to the reliance on many generated features. Among them, the SVM realized an F1-score of 73.63%, 2.39% and 5.23% higher than RF and DT, respectively. The neural models performed better than the traditional models, because they are capable of capturing more features by detecting the data relationships between different disease symptoms. Specifically, the RNN performed similarly with CNN, which achieved an excellent diagnosis effect with an F1-score of 80.57%. Thus, neural models are available for our task. Therefore, the CNN was adopted to train our network, and learn the internal representations between the original inputs and the final outputs in the DL.

Further, all the hyper parameters of our model were tuned via 5-fold cross-validation. The optimized hyper parameters are recorded in Table 3.

The experimental results show that our method achieved an overall F1-score of 86.77%, up by 7.7% and 17.85% from that of CNN and SVM, respectively. Thus, our model outperforms all the contrastive methods. Compared with the typical CNN, KG_CNN and TL_CNN improved the F1-score by 6.58% and 4.17%, respectively, reflecting the feasibility of diagnosing diseases through DL with the aid of knowledge graph and transfer learning.

Table 3. Optimized hyperparameters of our model

Hyperparameter	**Description**	**Value**
k	Dimension of word vector	50
d	Dimension of entity vector	50
h	Size of convolution kernel	2, 3, 4
l	Number of convolution kernels	150
dropout_rate	Dropout rate	0.75
learing_rate	Learning rate	0.2
batch_size	Batch size	16

Knowledge-based models

This subsection further discusses the effects of introducing experimental knowledge from medical knowledge graph of dairy cow into CNN. Table 4 compares the typical CNN (single CNN) model with our knowledge-based CNN models, which are entity_CNN (CNN with entity embedding), context_CNN (CNN with context embedding), and KG_CNN (CNN with entity and context embedding), respectively.

Table 4. Comparison of typical CNN and our knowledge-based models

Models	**Accuracy, (%)**	**Precision, (%)**	**Recall, (%)**	**F1, (%)**
CNN	81.22	80.93	80.21	80.57
entity_CNN	**84.05**	**83.73**	**83.34**	**83.53**
context_CNN	**83.24**	**83.11**	**82.18**	**82.64**
KG_CNN	**86.22**	**85.52**	**86.22**	**85.87**

As shown in Table 4, the knowledge-based CNNs, which were prepared under the knowledge extraction mechanism, outshined the typical CNN, revealing the prominent role of the knowledge extraction module in the experiment. The results demonstrate that our KG_CNN model realized an F1-score of 85.87%, better than that of the typical CNN. The entity_CNN achieved an F1-score of 83.53%, which improves the performance of CNN by 3.67%. Similarity, F1-score of the context_CNN model reached 82.64%, which is 2.57% higher than that of CNN. Moreover, the KG_CNN improved the performance of CNN by 6.58% in terms of F1-score. In addition, KG_CNN model outperformed entity_CNN and context_CNN by 2.8% and 3.91%, respectively.

It can be inferred from the above analysis that either entity_CNN or context_CNN can obtain a better performance than typical CNN, the reason is that the medical knowledge graph contains lots of disease and symptom entities. Once such knowledge is imported to the CNN, the network could learn the text representation from the rich semantic information, and give full consideration to the influence of different symptom features over disease classification, resulting in better classification performance. As for the better F1-score of our method, the good performance is mainly attributed to the entity and context vectors adopted in KG_CNN. These two parts are directly associated with each other. This structural information enhances the semantic representation of the medical records, and improves the ability of the model to obtain high-level text features. Therefore, it is feasible and effective to introduce structural knowledge of veterinary practices into disease diagnosis of dairy cow.

Table 5 compares our KG_CNNs represented by different knowledge graph embedding methods, including TransE [3], TransH [33], TransR [17], and TransD [15].

Table 5. Comparison between KG_CNNs represented by different knowledge graph embedding methods

Models	Accuracy, (%)	Precision, (%)	Recall, (%)	F1, (%)
TransE-KG_CNN	85.83	84.64	86.51	85.56
TransH-KG_CNN	84.97	84.10	85.13	84.61
TransR-KG_CNN	85.42	84.65	85.47	85.06
TransD-KG_CNN	**86.13**	**85.28**	**86.33**	**85.80**

As shown in Table 5, the KG_CNN trained by TransD achieved better performance than the model trained by other knowledge graph embedding methods, as evidenced by its high F1-score (85.80%). The reason is that TransD is more complex than TranE, TransH, and TransR in structure. The sophisticated structure can effectively capture the nonlinear relationships between structured knowledge, and reduce the semantic loss in vector quantization process. As a result, it is possible to acquire high-quality entities and entity context vectors. If these are imported to KG_CNN, the disease diagnosis will be much more effective.

Conclusions

Based on medical records, this paper proposes a dairy cow disease diagnosis method through DL driven by knowledge graph and transfer learning. Our method introduces the empirical knowledge of veterinarians to the CNN in the form of a knowledge graph, such that the network could learn more relevant features through the training on medical records. To further improve its performance, the disease diagnosis model was pretrained via transfer learning on the data generated from professional books. Experimental results demonstrate that our model is effective in clinical diagnosis of dairy cow. In the future, the authors will improve the proposed model with more complex architecture of the network and better medical knowledge.

Acknowledgements

This research was funded by the "Young Talents" Project of Northeast Agricultural University (Grant No. 17QC21), the National Key Research and Development (R&D) Program of China (Grant No. 2019YFE0125600), the China Agricultural Research System (Grant No. CARS-36), the University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province (Grant No. UNPYSCT-2020092), and the "Young Talents" Project of Northeast Agricultural University (Grant No. 17QC20).

References

1. Basavegowda H. S., G. Dagnew (2020). Deep Learning Approach for Microarray Cancer Data Classification, CAAI Transactions on Intelligence Technology, 5(1), 22-33.
2. Berredjem T., M. Benidir (2018). Bearing Faults Diagnosis Using Fuzzy Expert System Relying on an Improved Range Overlaps and Similarity Method, Expert Systems with Applications, 108, 134-142.
3. Bordes A., N. Usunier, A. Garcia-Duran, J. Weston, O. Yakhnenko (2013). Translating Embeddings for Modeling Multi-relational Data, Neural Inf Process Syst, 1-9.

4. Che W., Z. Li, T. Liu (2010). LTP: A Chinese Language Technology Platform, Coling 2010: Demonstrations, 13-16.
5. Chen Y., C. Y. Hsu, L. Liu, S. Yang (2012). Constructing a Nutrition Diagnosis Expert System, Expert Systems with Applications, 39(2), 2132-2156.
6. Dogantekin E., A. Dogantekin, D. Avci (2011). An Expert System Based on Generalized Discriminant Analysis and Wavelet Support Vector Machine for Diagnosis of Thyroid Diseases, Expert Systems with Applications, 38(1), 146-150.
7. Fricke P. M., J. O. Giordano, A. Valenza, G. Lopes Jr., M. C. Amundson, P. D. Carvalho (2014). Reproductive Performance of Lactating Dairy Cows Managed for First Service Using Timed Artificial Insemination with or without Detection of Estrus Using an Activity-monitoring System, Journal of Dairy Science, 97(5), 2771-2781.
8. Gao H., G. Jiang, X. Gao, J. Xiao, H. Wang (2019). An Equine Disease Diagnosis Expert System Based on Improved Reasoning of Evidence Credibility, Information Processing in Agriculture, 6(3), 414-423.
9. Grinter L. N., M. R. Campler, J. H. C. Costa (2019). Technical Note: Validation of a Behavior-monitoring Collar's Precision and Accuracy to Measure Rumination, Feeding, and Resting Time of Lactating Dairy Cows, Journal of Dairy Science, 102, 3487-3494.
10. Guo Y., J. Wang, H. Chen, G. Li, R. Huang, Y. Yuan, Y., T. Ahmad, S. Sun (2019). An Expert Rule-based Fault Diagnosis Strategy for Variable Refrigerant Flow air Conditioning Systems, Applied Thermal Engineering, 149, 1223-1235.
11. Gusterer E., P. Kanz, S. Krieger, V. Schweinzer, D. Süss, L. Lidauer, F. Kickinger, M. Ohlschuster, W. Auer, M. Drillich, M. Iwersen (2020). Sensor Technology to Support Herd Health Monitoring: Using Rumination Duration and Activity Measures as Unspecific Variables for the Early Detection of Dairy Cows with Health Deviations, Theriogenology, 157, 61-69.
12. Hamedan F., A. Orooji, H. Sanadgol, A. Sheikhtaheri (2020). Clinical Decision Support System to Predict Chronic Kidney Disease: A Fuzzy Expert System Approach, International Journal of Medical Informatics, 138, 104134.
13. Hu M., X. Chen, Y. Sun, X. Shen (2019). A Disease Prediction Model Based on Dynamic Sampling and Transfer Learning, Chinese Journal of Computers, 42(10), 2339-2354.
14. Hussain A., S. J. Lee, M. S. Choi, F. Brikci (2015). An Expert System for Acoustic Diagnosis of Power Circuit Breakers and On-load Tap Changers, Expert Systems with Applications, 42(24), 9426-9433.
15. Ji G., S. He, L. Xu, K. Liu, J. Zhao (2015). Knowledge Graph Embedding via Dynamic Mapping Matrix, Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Vol. 1, 687-696.
16. Kamphuis C., E. Frank, J. K. Burke, G. A. Verkerk, J. G. Jago (2013). Applying Additive Logistic Regression to Data Derived from Sensors Monitoring Behavioral and Physiological Characteristics of Dairy Cows to Detect Lameness, Journal of Dairy Science, 96(11), 7043-7053.
17. Lin H. L., Y. Liu, W. P. Wang, Y. L. Yue, Z. Lin (2017). Learning Entity and Relation Embeddings for Knowledge Resolution, Procedia Computer Science, 108, 345-354.
18. Liu B., Y. Xiao, Z. Hao (2018). A Selective Multiple Instance Transfer Learning Method for Text Categorization Problems, Knowledge-based Systems, 141, 178-187.
19. Ljubic B., S. Roychoudhury, X. H. Cao, M. Pavlovski, S. Obradovic, R. Nair, L. Glass, Z. Obradovic (2020). Influence of Medical Domain Knowledge on Deep Learning for Alzheimer's Disease Prediction, Comp Methods and Progr in Biomed, 197, 105765.
20. Løvendahl P., M. G. G. Chagunda (2010). On the Use of Physical Activity Monitoring for Estrus Detection in Dairy Cows, Journal of Dairy Science, 93(1), 249-259.

21. Luo A., S. Gao, Y. Xu (2018). Deep Semantic Match Model for Entity Linking Using Knowledge Graph and Text, Procedia Computer Science, 129, 110-114.
22. Madureira A. M. L., B. F. Silper, T. A. Burnett, L. Polsky, L. H. Cruppe, D. M. Veira, J. L. M. Vasconcelos, R. L. A. Cerri (2015). Factors Affecting Expression of Estrus Measured by Activity Monitors and Conception Risk of Lactating Dairy Cows, Journal of Dairy Science, 98(10), 7003-7014.
23. Maini E., B. Venkateswarlu, B. Maini, D. Marwaha (2021). Machine Learning-based Heart Disease Prediction System for Indian Population: An Exploratory Study Done in South India, Medical Journal Armed Forces India, https://doi.org/10.1016/j.mjafi.2020.10.013.
24. Nugroho A. (2017). Mobile Expert System Using Fuzzy Tsukamoto for Diagnosing Cattle Disease, Procedia Computer Science, 116, 27-36.
25. Nusai C., S. Cheechang, S., Chaiphech, G. Thanimkan (2015). Swine-vet: A Web-based Expert System of Swine Disease Diagnosis, Procedia Computer Science, 63, 366-375.
26. Pham T., T. Tran, D. Phung, S. Venkatesh (2017). Predicting Healthcare Trajectories from Medical Records: A Deep Learning Approach, J Biomed Inf, 69, 218-229.
27. Piette D., T. Norton, V. Exadaktylos, D. Berckmans (2020). Individualised Automated Lameness Detection in Dairy Cows and the Impact of Historical Window Length on Algorithm Performance, Animal, 14(2), 409-417.
28. Reiter S., G. Sattlecker, L. Lidauer, F. Kickinger, M. Öhlschuster, W. Auer, V. Schweinzer, D. Klein-Jöbstl, M. Drillich, M. Iwersen (2018). Evaluation of an Ear-tag-based Accelerometer for Monitoring Rumination in Dairy Cows, J Dairy Sci, 101, 3398-3411.
29. Sajadi N. A., S. Borzouei, H. Mahjub, M. Farhadian (2019). Diagnosis of Hypothyroidism Using a Fuzzy Rule-based Expert System, Clin Epidem Global Health, 7(4), 519-524.
30. Shang X. Z., M. Han, S. F. Wang, T. Jia, G. Y. Xu (2020). A Skin Disease Diagnosis Method Combing Transfer Learning and Neural Networks, CAAI Transactions on Intelligent Systems, 15, 452-459.
31. Zhao J., S. Gu, A. McDermaid (2019). Predicting Outcomes of Chronic Kidney Disease from EMR Data Based on Random Forest Regression, Mathem Biosc, 310, 24-30.
32. Zheng Y., B. K. Iwana, S. Uchida (2019). Mining the Displacement of Max-pooling for Text Recognition, Pattern Recognition, 93, 558-569.
33. Zhu Q. N., X. F. Zhou, P. Zhang, Y. Shi (2019). A Neural Translating General Hyperplane for Knowledge Graph Embedding, Journal of Computational Science, 30, 108-117.

Meng Gao, Ph.D.
E-mail: gaomeng@neau.edu.cn

Meng Gao is a Lecturer in the College of Electrical Engineering and Information, Northeast Agricultural University, Harbin, China. She obtained her Ph.D. from College of Information and Computer Engineering, Northeast Forestry University, Harbin, China. Her current research includes livestock big data, knowledge graph, machine learning and decision support system.

Haodong Wang, M.Sc. Student
E-mail: whaodong1997@163.com

Haodong Wang is a Postgraduate at the College of Electrical Engineering and Information, Northeast Agricultural University, Harbin, China. He obtained his B.Sc. degree in School of Computer Science and Technology from Shandong Jianzhu University, Jinan, China. His research interests include data mining, natural language processing and knowledge graph.

Prof. Weizheng Shen, Ph.D.
E-mail: wzshen@neau.edu.cn

Weizheng Shen is a Professor at the College of Electrical Engineering and Information, Northeast Agricultural University, Harbin, China. He obtained his Ph.D. degree from the same university. His current research includes livestock big data, intelligent perception and computing, and efficient and healthy breeding facilities. He has authored/co-authored over 50 research articles.

Prof. Zhongbin Su, Ph.D.
E-mail: suzb001@163.com

Zhongbin Su is a Professor at the College of Electrical Engineering and Information, Northeast Agricultural University, Harbin, China. He obtained his Ph.D. degree from the School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China. His current research includes agricultural big data, intelligent monitoring of crops, and decision support system. He has authored/ co-authored over 50 research articles.

Huihuan Liu, B.Sc.
E-mail: liuhuihuan@sina.com

Huihuan Liu is a Senior Veterinarian of Dairy Association of Heilongjiang Province, China. She received her B.Sc. degree in College of Veterinary Medicine from Northeast Agricultural University, Harbin, China. She has worked as a competent veterinary in a dairy cow farm for 8 years. Previously, she is occupied in the promotion of dairy farming technology.

Yanling Yin, Ph.D.
E-mail: yinyanling@neau.edu.cn

Yanling Yin is a Lecturer in the College of Electrical Engineering and Information, Northeast Agricultural University, Harbin, China. She obtained her Ph.D. in College of Underwater Acoustic Engineering from Harbin Engineering University, Harbin, China. Her current research includes environmental control of livestock housing, intelligent perception and computing.

Prof. Yonggen Zhang, Ph.D.
E-mail: zhangyonggen@sina.com

Yonggen Zhang is a Professor at the College of Animal Sciences and Technology, Northeast Agricultural University, Harbin, China. He obtained his Ph.D. degree from the same university. His current research includes development of unconventional feed recourses for ruminants and healthy breeding of dairy cows. He has authored/co-authored over 300 research articles.

Assoc. Prof. Yi Zhang
E-mail: chrisrang@163.com

Yi Zhang is an Associate Professor in College of Computer Science and Electrical Engineering, East University of Heilongjiang, Harbin, China. He obtained his M.Sc. degree in Software College from Heilongjiang University, Harbin, China. His research interests include livestock big data, applications of Internet of Things, data visualization.