

## Pet dog disease pre-diagnosis system for caregiver with possibilistic C-Means clustering and disease database

Kwang Baek Kim<sup>1</sup>, Doo Heon Song<sup>2</sup>

<sup>1</sup>Division of Computer and Information Engineering, Silla University, Korea

<sup>2</sup>Department of Computer Games, Yong-In SongDam College, Korea

---

### Article Info

#### Article history:

Received Jan 16, 2020

Revised Mar 14, 2020

Accepted Apr 11, 2020

---

#### Keywords:

Disease diagnosis

Health monitoring

Machine learning

Pet dog

Possibilistic C-Means

---

### ABSTRACT

While the population of pet dogs and veterinary clinics are increasing, there is no reliable and useful software for pet owners/caregivers who have limited knowledge on the pet diseases. In this paper, we propose a pre-diagnosis system working on the mobile platform that the pet owner can take a pre-diagnosis from his/her observation of pet dog's abnormality. Technically, the system needs a reliable databases for disease-symptom association thus we provide it based on the textbook and encyclopedia. Then, we apply Possibilistic C-Means algorithm that is an unsupervised machine learning algorithm to form the connections between disease and symptoms from database. The system outputs five most probable diseases from the observed symptoms of pet dog. The utility of this system is to alert the owner's attention on the pet dog's abnormal behavior and try to find the diseases as soon as possible.

*Copyright © 2020 Institute of Advanced Engineering and Science.  
All rights reserved.*

---

### Corresponding Author:

Kwang Baek Kim,

Division of Computer and Information Engineering,

Silla University, Busan 46958, Korea

Email: gbkim@silla.ac.kr

---

## 1. INTRODUCTION

The population of pets kept in households have increased worldwide recently. It is reported that 43% of families own dogs or cats in Italy [1]. In Korea, dogs become almost family members of their owners as living in the same house and pet shops and veterinary clinics are increasing in the big cities. Problems arise when pet dogs feel uncomfortable but owners who have limited knowledge about canine health care do not catch the incident or give inappropriate care. Usually a pet dog gives a sign to its owner by expressing unusual behavior or by the change of its body when its health is at risk or having a disease. However, without deep knowledge about the pet dog's disease, owners tend to neglect such signs but depend on the regular check by veterinarians only to make the situation worse [2]. Some owners try to treat these infected dogs on their own and make the status worse and make dogs behaving fierce or tired and inactive. Knowing about common dog diseases and being aware of appropriate treatment can better help the owner in taking care of their dogs [3].

Effective canine health surveillance provides information that supports disorder prioritization, improved disorder management, and focused clinical research [4]. Optimal data sources for canine disease surveillance require representativeness of the general population, a well-defined denominator population, validity of disorder diagnosis data and sustainability. Limitations in these areas present as selection bias, misclassification bias and discontinuation of the system respectively [5]. Primary data that are collected specifically for the research offers better control of the types and quality of the data and may be easier to validate. However, secondary data that have been pre-collected by a third party offer more efficient collection, reduced costs, larger sample size, better representativeness and reduced bias [4]. Secondary sources of health data for dogs include pet insurance, referral practice and primary-care practice records [5-7] and evaluated as more reliable one [4]. Similarly, the World Health Organization (WHO) International

Classification of Diseases (ICD-10) for human diseases has been adapted to canine and feline diseases to standardize the diagnostic nomenclature [1].

Thus, we need a pre-diagnosis system [8, 9] for pet dog owners/caregivers who have limited knowledge on the treatment policy when the incident happens. Such pre-diagnosis system can accept symptoms that the caregiver can observe as input and produces the list of probable diseases names and possible treatments as output. The main purpose of such pre-diagnosis system is not to replace veterinarian but to alert the caregiver and find possible disease in its early stage and take appropriate treatment from veterinary clinics afterwards. For example, canine cataract can be found from rough cellular phone images taken by caregiver from a pre-diagnosis system [10-12].

There have been several researches to make computer aided decision making systems for cattles' health monitoring as the form of expert system [3, 13-17]. Those systems can handle a limited range of diseases for a certain cattle such as dog [3], horse [13], cow [14], fish [15], swine [16], and general cattle [17]. Usually their data collection method is primary and only can handle less than 10 related diseases with traditional certainty factor [13-16] or fuzzy logic [17]. Thus the utility of such system is very limited and possibly unreliable or subjective.

In this paper, we propose a pre-diagnosis system that can work in the mobile platform using a machine learning technique called Possibilistic C-Means (PCM) clustering algorithm [18]. PCM is an unsupervised learning algorithm thus we do not have to be influenced by a small number of experts' opinions with primary data collection. PCM has rich theoretical background [18] and many successful applications in engineering [19, 20] and medical domain [21, 22]. In data collection, we make symptom-disease association database based on the encyclopedia style textbooks [23, 24] as we did in traditional medicine database [25]. Thus, unlike other researches with small number of expert-generated rules, we have hundreds of symptom-disease associations from textbook and let the unsupervised learner make the diagnosis from pet caregiver's symptom input.

**2. METHOD**

**2.1. Symptom-Disease Database Structure**

For the pet dogs' disease-symptom data, we used two books [23, 24] to collect important/identifiable diseases by observation and related symptoms. Total 100 diseases and 413 symptoms are collected as symptom-diseases database with respect to 5 pet dog body parts that those symptoms occur - whole body, leg, face, skin, and suspicious internal uncomfortableness. Those symptoms are associated with diseases and verified by veterinarians. Some representative symptoms associated with a certain disease also creates other queries to user for further considerations.

Obtained symptoms and corresponding diseases are associated with different strength that will be learned by our PCM learning engine. Our database consists of three main tables as following;

- a) Disease (ID, DiseaseCode, SymptomCode, Description)
- b) Symptom (SymptomID, BodyPart, Description)
- c) LearnedResult (ClusterID, InputNeuron, Strength)

SymptomCode of Disease table may contain multiple SymptomIDs of Symptom table. Since usually a disease has multiple symptoms. The learning result table stores the connection strength after learning between disease and symptom as shown arrows in Figure 1.

The field Symptomcode consists of code-code-...-code and the delimiter '-' is removed when the query is processed. Each symptom has the description field that explains what the symptom means and how the treatment should be given. The example database is as shown in Table 1 and Table 2.

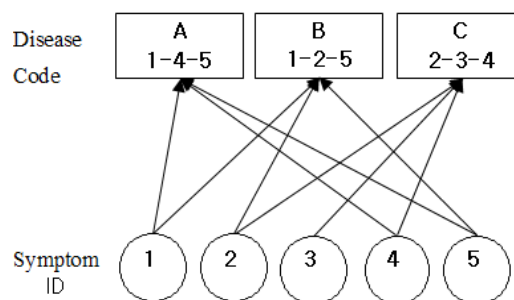


Figure 1. Symptom-Disease association

Table 1. Disease table example

ID	DISEASE	SYMPTOM CODES	DESCRIPTION
1	ascariasis	2-3-4-5-62-70	Expression
2	trichuriasis	1-2-6-62-70-71-80	Expression
3	.	.	.
100	prostatomegaly	2-70-80-171	Expression

Table 2. Symptom table example

ID	PART	SYMPTOM
1	Whole Body	anemia
2	Whole Body	inappetence
3	.	.
413	Inside	Heartburn

**2.2. Learning by PCM for diagnosis**

Then, the next step is applying unsupervised learning method to compute the degree of association between disease and symptoms as shown in Figure 1. Our model is based on the standard PCM suggested in [18]. In PCM, the generated component corresponds to a dense region in the dataset. PCM is known to be more robust in the case of noisy data. The PCM membership degree refers to the degree of ‘typicality’ between data and clusters [21]. The PCM is based on the relaxation of the probabilistic constraint in order to interpret in a possibilistic sense of the membership function or degree of typicality.

Then we call  $U_p = (u_{ij})$  a *possibilistic cluster partition* of  $X$  if

$$\sum_{j=1}^n u_{ij} > 0, \forall i \in \{1, \dots, c\} \tag{1}$$

The  $u_{ij} \in [0, 1]$  are interpreted as degree of typicality of the datum  $x_j$  to cluster  $i$  and  $u_{ij}$  for  $x_j$  resembles the possibility of being a member of corresponding cluster.

The objective function in PCM clustering can be calculated as follows

$$J_{PCM} = \sum_{k=1}^n \sum_{i=1}^c t_{ik} d_{ik}^2 + \sum_{i=1}^c \sum_{k=1}^n (1 - t_{ik})^m \tag{2}$$

where  $d_{ik}$  denotes the average distance between points in the same group in this paper and there are  $c$  clusters and  $n$  symptoms. The term  $t_{ik}$  denotes the typicality of the  $k^{th}$  data point belongs to the  $i^{th}$  cluster and  $m$  is the empirical constant set to 2.0 in this paper. The first term of (1) tries to minimize the distance between the data point and the cluster center and the second term is a penalty term to avoid having a trivial solution.

$$t_{ik} = \frac{1}{1 + \left(\frac{d_{ik}^2}{\delta_i}\right)^{1/m-1}} \tag{3}$$

The update formula for membership degree is defined as (2) where  $\delta_i$  denotes the area of the  $i^{th}$  cluster computed by (3).

$$\delta_i = \frac{\sum_{k=1}^n t_{ik}^m d_{ik}^2}{\sum_{k=1}^n t_{ik}^m} \tag{4}$$

As shown in (1) can be rewritten as (4) that describes the object function as the sum of  $c$  independent values.

$$J_{PCM} = \sum_{i=1}^c \left( \sum_{k=1}^n t_{ik}^m d_{ik}^2 + \sum_{k=1}^n (1 - t_{ik})^m \right) \tag{5}$$

This process continues until the update rule changes the typicality no more than the predefined threshold.

The result of PCM is the association of each symptoms to the disease clusters with typicality. The diagnosis of pet dog health monitoring is done in the following way.

- a. Input abnormal body part (one of whole body, leg, face, skin, and internal).
- b. Choose observed symptoms of the pet dog of the body part.
- c. The system explores the symptom-disease association space learnt by PCM and take five related diseases with highest typicality.

**3. IMPLEMENTATION RESULT**

The proposed method is implemented in Android Studio Version 5.0.0 and Android Simulator SDK 22 (lollipop) with Intel(R) Dual Core (TM) i5-66005U CPU @ 3.3 GHz and 8 GB RAM PC. The user interface of the system is as shown in Figure 2.

Then the caregiver select observed symptoms with respect to the body part chosen from Figure 2 (a) as shown in Figure 2 (b). Then the system gives most probable (highest typicality) disease based on the input as shown in Figure 3 with disease description.

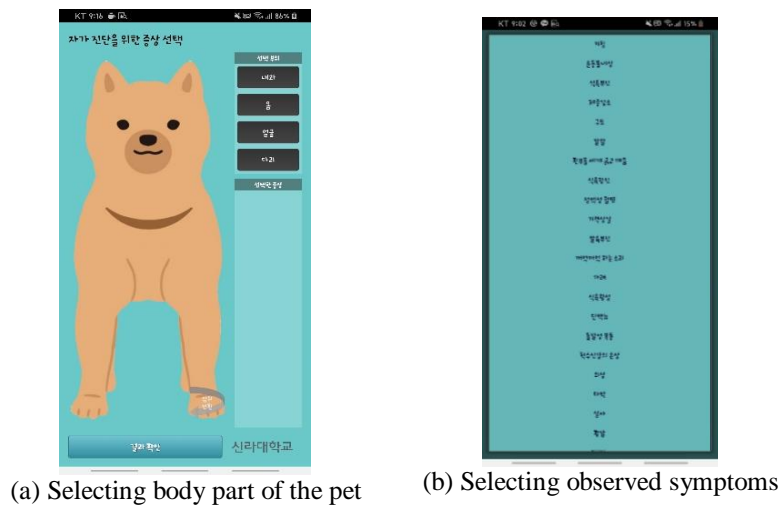


Figure 2. User interfaces

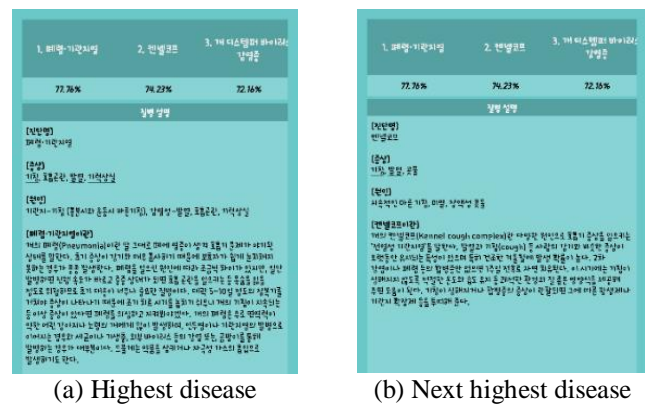


Figure 3. User interface of system output

**4. CONCLUSION**

In this paper, we propose a pre-diagnosis system working on the mobile platform for pet dog owners/caregivers. While rapidly increasing pet dogs around human being worldwide, there is no systematic reliable pet dog disease databases and health monitoring systems for pet owners. The traditional expert system approach might not work since such primary data collection system can not cover various diseases

that the owner might observe but have no knowledge what to do. The system we propose consists of the database collected from encyclopedia and textbook and the unsupervised learner called PCM that forms a number of disease clusters associated with various related symptoms with respect to the body parts of the pet and the degree of typicality. The implemented software takes observed abnormal symptom from the owner as input and the PCM learner explores the symptom-disease associations and outputs 5 most probable diseases with description. However, the role of this system is not replacing veterinarian's diagnosis but stimulating owner's attention to pet dog's abnormality in nature. There are several similar diseases with very similar symptoms that an ad-hoc observation cannot see the difference. Also, some diseases have different symptoms with respect to the progress of the disease. In those cases, veterinarian's care is "must".

## REFERENCES

- [1] M. Martini, *et al.*, "A surveillance system for diseases of companion animals in the Veneto region (Italy)," *Scientific and Technical Review of the Office International des Epizooties*, vol. 36, no. 3, pp. 1007-1014, 2017.
- [2] K. B. Kim, *et al.*, "Machine intelligence can guide pet dog health pre-diagnosis for casual owner: a neural network approach," *International Journal of Bio-Science and Bio-Technology*, vol. 6, no. 2, pp. 83-90, 2014.
- [3] M. Y. Munirah, *et al.*, "Design and development of online dog diseases diagnosing system," *International Journal of Information and Education Technology*, vol. 6, no. 11, pp. 913-916, 2016.
- [4] D. G. O'Neill, *et al.*, "Approaches to canine health surveillance," *Canine genetics and epidemiology*, vol. 1, no. 1, pp. 2, 2014. [online]
- [5] M. G. Doherr and L. Audigé, "Monitoring and surveillance for rare health-related events, a review from the veterinary perspective," *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 2001, vol. 356, pp. 1097-1106.
- [6] H. T. Sorensen, *et al.*, "A framework for evaluation of secondary data sources for epidemiological research," *Int J Epidemiol*, vol. 25, no. 2, pp. 435-442, 1996.
- [7] A. Egenvall, *et al.*, "Insurance data for research in companion animals, benefits and limitations," *Acta Vet Scand*, vol. 51, no. 1, pp. 42, 2009.
- [8] Nilesh Kulkarni, "Support vector machine based alzheimer's disease diagnosis using synchrony features," *International Journal of Informatics and Communication Technology*, vol. 9, no. 1, pp. 57-62, 2020.
- [9] V. M. Nisha and L. Jegathanan, "A symmetry based anomaly detection in brain using cellular automata for computer aided diagnosis," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 1, pp. 471-477, 2019.
- [10] J. E. Buring, "Primary data collection, what should well-trained epidemiology doctoral students be able to do?," *Epidemiolog*, vol. 19, no. 2, pp. 347-349, 2008.
- [11] K. B. Kim, "Extraction of canine cataract object for developing handy pre-diagnostic tool with fuzzy stretching and art2 learning," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 16, no. 1, pp. 21-26, 2016.
- [12] K. B. Kim and D. H. Song, "Intelligent Automatic Extraction of Canine Cataract Object with Dynamic Controlled Fuzzy C-Means based Quantization," *International Journal of Electrical & Computer Engineering*, vol. 8, no. 2, pp. 666-672, 2018.
- [13] H. Qin, *et al.*, "Horse-Expert: An aided expert system for diagnosing horse diseases," *Polish journal of veterinary sciences*, vol. 19, no. 4, pp. 907-915, 2016.
- [14] L. Rong and D. Li, "A web based expert system for milch cow disease diagnosis system in China," *International Conference on Computer and Computing Technologies in Agriculture*, Springer US, pp. 1441-1445, 2007.
- [15] D. Li, *et al.*, "Fish-Expert: a web-based expert system for fish disease diagnosis," *Expert systems with Applications*, vol. 23, no. 3, pp. 311-320, 2002.
- [16] C. Nusai, *et al.*, "Swine-Vet: a Web-based expert system of swine disease diagnosis," *Procedia computer science*, vol. 63, pp. 366-375, 2015.
- [17] A. Nugroho, *et al.*, "Mobile expert system using fuzzy tsukamoto for diagnosing cattle disease," *Procedia computer science*, vol. 116, pp. 27-36, 2017.
- [18] N. R. Pal, *et al.*, "A possibilistic fuzzy c-means clustering algorithm," *IEEE transactions on fuzzy systems*, vol. 13, no. 4, pp. 517-530, 2005.
- [19] A. Bhat, "Possibility fuzzy c-means clustering for expression invariant face recognition," *Int. J. Cybern. Inf.*, vol. 3, no. 2, pp. 35-45, 2014.
- [20] A. Suryadibrata and K. B. Kim, "Ganglion Cyst Region Extraction from Ultrasound Images Using Possibilistic C-Means Clustering Method," *Journal of information and communication convergence engineering*, vol. 15, no. 1, pp. 49-52, 2017.
- [21] S. R. Kannan, *et al.*, "Effective fuzzy possibilistic c-means: an analyzing cancer medical database," *Soft Computing*, vol. 21, no. 11, pp. 2835-2845, 2017.
- [22] J. Aparajeta, *et al.*, "Modified possibilistic fuzzy C-means algorithms for segmentation of magnetic resonance image," *Applied Soft Computing*, vol. 41, pp. 104-119, 2016.
- [23] Korea Society of Veterinary Science, "The encyclopedia of veterinary science," *Korea Society of Veterinary Science*, 2003.
- [24] N. J. Kim, *Pet Dog Diseases*, 21 century publishing Co. 2013.
- [25] K. B. Kim, *et al.*, "Developing Intelligent Health Diagnosis System for Korean Traditional Medicine: Database Construction and Neural Inferences," *International Journal of Bio-Science and Bio-Technology*, vol. 6, no. 3, pp. 1-8, 2014.

**BIOGRAPHIES OF AUTHORS**

**Kwang Baek Kim** received his M.S. and Ph.D. degrees from the Department of Computer Science, Pusan National University, Busan, Korea, in 1993 and 1999, respectively. From 1997 to the present, he is a professor at the Department of Computer Engineering, Silla University, Korea. He is currently an associate editor for Journal of Intelligence and Information Systems and The Journal of Information and Communication Convergence Engineering. His research interests include fuzzy clustering and fuzzy control system, data mining, image processing, computer vision and bioinformatics.



**Doo Heon Song** received his B.S. degree in Statistics & Computer Science from Seoul National University, Korea and M.S. degree in Computer Science from the Korea Advanced Institute of Science and Technology in 1983. He received his Ph.D. Certificate in Computer Science from the University of California at Irvine in 1994. He has been a professor at the Department of Computer Games, SongDam College, Korea, since 1997. He has served as an associate editor for Journal of Multimedia Signal Processing and Information Hiding and The Journal of Information and Communication Convergence Engineering. His research topics include artificial intelligence, machine learning and video game design & culture.