Journal of Engineering and Applied Sciences 13 (20): 8591-8597, 2018
ISSN: 1816-949X

The Data Mining Reliability for Melanoma Disease Diagnosis

Hussein M. Haglan and Akeel Sh. Mahmoud
Computer Center, University of Anbar, Anbar, Ramadi, Iraq

Abstract: Data mining methods are the amount of actual data are used to study these data to forecast entire some data to support a decision-making in a problem-solving. A data mining is very beneficial to study any disease parameters to support the decision development and specify the disease and details. In the proposed present studies, using the real algorithms of data mining methods to support various healthcare fields and accepted a correct decision about the diagnosis of melanoma disease and specify the risk reasons for this disease to support decision process. In this study, a data-mining technique of melanoma disease forecast using a mixed scheme of Backpropagation-Neural Network (Bp-NN) and Genetic Algorithms (GA) has been introduced. According the outcomes, it has been seen that a mixed model forecast melanoma disease with nearly 95% accuracy. Additionally, the tested samples of entities share the same risk factors a symptom. Data mining depends on these symptoms and parameters to detect melanoma disease.

Key words: Melanoma sisease, Decision Support (DS), data mining, Genetic Algorithm (GA), Backpropagation Neural Network (Bp-NN), beneficial

INTRODUCTION

Recently, melanoma disease (also, called Melanoma Malignant Tumor(MMT)) is the main reason of death in the world. The Organization of World Health (WHO) introduced a report approved that 16 million cases of death happen around the world due to melanoma disease. The rate of death has increasing over 75% by growing the causes (hormonal factors, family history 5%, sex factors, pathological factors and Degeneration of melanocytes) (WHO., 2011; Jamison *et al*., 2006)

The early, accurate prognosis of the disease should reduce the risks of melanoma disease. The detections begin with symptoms based on multi X-ray of the melanoma disease to detect and diagnose the disease exactly and which phase became. The complete healthcare consist of a large data about any disease, particularly melanoma disease. This healthcare contains of hidden data, this data is very useful to creation an accurate making actual decision particularly melanoma patients. There are a large amount of healthcare complicated data belong to hospitals sources, disease diagnosis, a full patient records as well as effective medical treatments and etc. These large, relevant data can be considered as the key to analyzing a large amount of data and then deriving knowledge from these data in order to make the appropriate and effective decision. A mixed method was developed using different and effective techniques to improve the accuracy of the disease diagnosis and the time it took for diagnosis (Bakpo and Kabari, 2011).

Literature review: To improve patient efficiency and age, most research is carried out in the field of disease detection by proposing new and up-to-date methods of effective health care.

In planned to presented data mining and its several techniques and a survey of the available works on medical data mining. We highlight mainly on the application of data mining on skin diseases (Barati *et al*., 2011). In proposed a cancer prognosis system depend on data mining technology. This technology approximations the risk of the breast cancer in the earlier stage. This scheme is authenticated by comparing its prognosis outcomes with patient's prior medical information and it was examined by using weka system (Priyanga and Prakasam, 2017). In presented a study that summarizes different technical papers on medical diagnosis and prediction. It has also, been focused on present paper being implemented by using the data mining methods to improve the disease (s) predicting process. This study provides future trends of current methods of KDD, using data mining tools for healthcare (Kaur and Bawa, 2015). In presented an improvement of a skin diseases diagnosis scheme which permits user to recognize diseases of the human skin and to medical counseling or medical treatments in a very short time period. For this reason, the user will have to upload an skin image disease to the system and answer questions depend on their skin condition or indications (Amarathunga *et al*., 2015). In presented an study features different Data Mining methods such as classification, clustering as sociation

Correspomding Author: Hussein M. Haglan, Computer Center, University of Anbar, Anbar, Ramadi, Iraq

and also focused related work to analyze and prognosis human disease (Patel and Patel, 2016). In focused in this study on the major clinical and histopathological attributes influences on psoriasis disease of human body. This study creates the relationship between the input and the response features for enhance disease diagnosis in medical field. The Response Surface Methodology (RSM) is make used for improve a relationship between input attributes of skin disease and prognosis the psoriasis patients with the help of independent and dependent variables (Sudha *et al*., 2017). In used a system based on the decision tree for mining and processing image data. This scheme will be used for classification of human skin diseases. The authors will try to use the decision tree and DIP (Digital Image Processing) principals to detect skin diseases by using some features that found in the skin digital image (Kadhim, 2017). In also, presented an effective model where 40 digital images are collected from AOCD unit database and another 40 digital images from MIT unit database. These proposed model compared one various established feature. The researcher also, experimented with various classifier architectures. The characteristic sets and most of the classifier architectures tested, provide a similar performance. To enhance the performance the outcomes of the experiments show that this model has much faster training and testing times than the widely used multi SVM methods (Parvin and Jafar, 2017). In constructed a diagnosis system depend on the methods of data mining and image processing. The proposed model will capture an image through the smartphone camera. The system have a big benefit to the dermatologists as a prescreening system for early diagnosis in situations where the dermoscopes are not reachable. The model will compare the acquired image with training dataset using image processing techniques and decides whether a skin suffers from diseases or not using decision tree (Gound *et al*., 2018).

MATERIALS AND METHODS

System techniques and resources: Therapeutic diagnosis procedure start when the patient with the specialists and offers a range of symptoms. The doctor asks the patient many different questions for the purpose of identifying and diagnosing severe cases of the disease. The data collected on the patient includes the patient's previous health conditions as well as his or her living conditions. After full physical examination of the patient, continuous monitoring (24/7) of the patient is performed before proper medical treatment is given. The images taken by dermoscopy have been ranked and classified to be suitable for our system (Fig. 1).

Fig. 1: Various kinds of dermoscopy devices

Data mining based technique: Data mining are a traditional methods to merge analyzing data process with complicated algorithms to extract ideal information, benefit from among a large amount of data is used; these data may be use later in the expectation an event in the future. Usually, these data is classified into three main sections:

- A data, raw image data without any test or analyze is
- An Information analyzed data and draw some simple information
- Information has been shown in more complex ways. If this information are added to the analyst, then, the experience becomes knowledge (Kumar *et al*., 2011)

Data mining levels

Level-a: Initial analysis and conversion-preprocessing and transformation: are in that stage different raw data is converted to the form and to the standard format in the subsequent phase of analysis, the longer this stage is the most time-consuming and effort.

Level-b: The Initial analysis and conversion-the preprocessing and transformation: at this stage, the various raw data are converted to the standard format the subsequent analysis stage. If this stage is longer then more time consuming and effortless it is (Bakpo and Kabari, 2011).

The algorithms of data mining: The algorithms of data mining (researcher called machine learning) is a set of calculations that is lead to generate a data model. In data mining, the algorithm analyzes the acquired data and seeking for exact patterns to generate a required model. The data mining algorithms can be categories into the following classifications.

Classification: Used to determine the compatibility of new hypotheses for any of the predefined categories.

Regression: to measure the relationship between various variables, they forecast the value of any variable based on the value of additional variable.

Clustering: they looking for a way to categorize information into similar blocks in the properties of each block are similar in their properties from the other cluster.

Sequential pattern discovery: Is the detection of a recurring pattern in a specific order as the obtaining of cold medicines followed by some foods helps to reduce the disease (Munnoli and Bapat, 2013; Bullinaria, 2004).

Disease description

Melanoma disease: It is also called malignant melanoma also known as malignant melanoma, a type of cancer that develops from cells containing the pigmented-melanin-responsible color of the skin-known as the melanocytic cell (Bullinaria, 2004). Melanoma usually occurs in the skin but rarely occurs in the mouth, intestine, eye (WHO., 2011; Jemison *et al*., 2006) or any other parts of the body. Tumor cell cancer is less common than other types of skin cancers but it is more serious and is attributed to the majority (75%) of skin cancer-related deaths. About 61,000 cases of melanoma are diagnosed every year throughout the world. The incidence of melanoma is higher in women than in men and in light-skinned people who live in sunny areas with high rates in Australia, New Zealand, North America and northern Europe, according to the World Health Organization. In its causes 48,000 deaths per year to pigment cell carcinoma (WHO., 2011).

Signs and symptoms of melanoma disease: Early signs of melanoma are changes in the shape and color of lumps or in the case of nodal melanoma showing new lumps anywhere on the skin (Kaur and Bawa *et al*., 2015). In later stages appear itching or may ulcers lump or bleed. Early signs of itching are summarized as follows:

- Asymmetry
- Irregular shelves
- Coloration
- Evolve over time
- High on the skin surface
- Solid, compact texture
- Increasing

The melanoma diagnosis: Visual examination is the most common diagnostic method for detecting melanoma (Kaur and Bawa *et al*., 2015). Usually irregular clusters are treated in color, size and shape as candidates for tumors. In order to detecting melanin tumors (and increasing survival rates), it is advisable to learn how to identify them to examine changes that occur regularly (shape, size, color, itching or bleeding) and consult a qualified doctor (Goldberg, 2002; Michel and Charles, 1999) (Fig. 2).

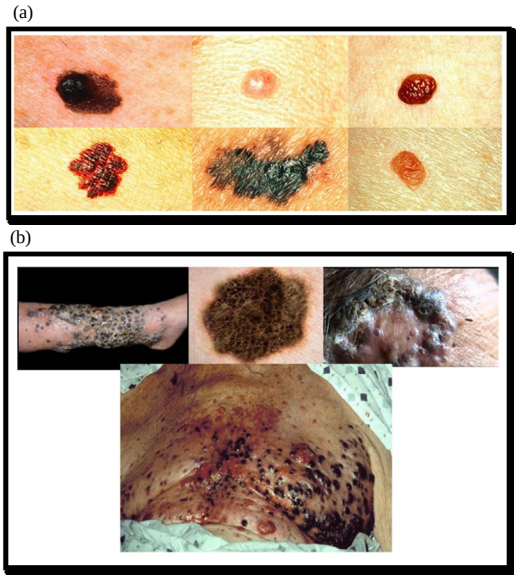

Fig. 2: Different cases of melanoma: a) Simple cases and b) Worst cases

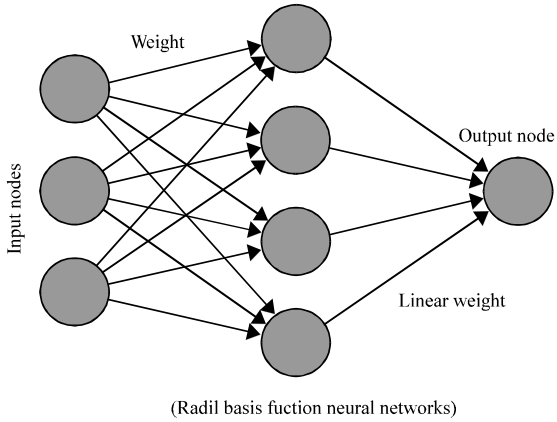

Fig. 3 :The main structure of Bp-NN

Artificial Neural Networks (ANN)

Backpropagation Neural Networks (Bp-NN): Bp-NN as shown in Fig. 3 is a supervised learning technique used to training artificial neural networks according Eq. 1. This technique was described by the scientist Rumelhart in 1974 and further enhancement and developed by the scientists Rumelhart, Hinton and Williams in 1986. It is most beneficial for feed-forward networks. The abbreviation term refer to "Forward errors propagation":

$$net_{pj}^{h} = \sum_{i=1}^{N} w_{ji}^{h} x_{pi} + \varnothing_{j}^{h} \quad (1)$$

The following description is the basic Bp-NN action (procedure) for network training:

- Apply an vector's input to the network and compute the matching output parameter
- Matching the real outputs with the accurate outputs and compute the error measure
- Specify in any direction (- or +) to change each weight in order to minimize the error
- Calculate the amount of change for each weight
- Apply the corrections to the weights
- Repeat elements from (Eq. 1-5) with all vector training until the error for all vectors in the training set is reduced to an acceptable value

Genetic Algorithm (GA) based proposed system: The algorithms optimization and machine learning are called GA. These algorithms based on many features of biological evolution. The GA requires 5 parameters: A method for encoding answers for the issue on chromosomes:

- An evaluation function which returns a rating for each chromosome given to it
- A way of initializing the population of chromosomes
- Operators that might be connected to parents when they imitate to adjust their genetic composition
- Parameter settings for the algorithm, the operators and so on (Goldberg, 2002)

Given these five components, a (GA) operates according to the following two steps:

- Initialize the population by using the procedure of initialization and compute each initial population member
- Reproduce until meet the stop criteria. Reproduction comprises of iterations of the accompanying three stages
- Select one or more parents to reproduce. Selection is stochastic; however, the individuals with the highest estimations are desired in the selection
- Choose a genetic operator and apply it to the parent
- Children estimation and accumulate into a generation. After accumulating sufficient individuals, put these individuals into the population, replacing the worst current members of the population (Michel and Charles, 1999)

Melanoma detection hybrid system: The proposed system is based on historical databases of patients with melanoma disease where this algorithm relies on the development of decision support for the prognosis of melanoma disease. This system depends on main medical parameters such as: (Size increasing, color change, itch and bleeding). The proposed Bp-NN contains an input layer to obtain an input signal of melanoma case and output layer to produce the preferred output to be inputted to the GA. The input layer contained of 3 neurons, the hidden layer had 25 neurons and the output layer had 5 neurons. The input layers (input values) are concerned with the information of patient such as (name, age and sex). The hidden layer called (middle layers or hidden values), these parameters are related with uses the patient data to make diagnoses. The output layer is called (output values), this layer produces the outputs (the values processing).

The Bp-NN learning begin by putting an melanoma image as blocks, each block (iteration) contains a matrix of 8×8 elements (PE) by multi iterations to proposed net. The first iteration puts into the net as an input block and applying feedforward NN. First input iteration compares with preferred output, if there is an error, then adjust the weights of each node by applying Bp-NN for the same iteration until arriving to same desired output, at this time can pick this iteration as a good population according to the map described in Fig. 4.

In the GA, the thresholding is a process to the process of natural selection the beginning, a small number of chromosomes may live. The new population would be produced, to find some chromosomes that permit the test. In order to produce parent's offspring, the threshold permits some of the chromosomes to continue when the threshold has value more than the cost value of chromosomes. Most of the chromosomes will live providing the threshold is not changed in the next generations.

Learning begins with Bp-NN, placing melanoma images as a set of so-called blocks (iterations). Each iteration consists of 8×8 matrix elements (process elements) for several iterations in the network. The first iteration enters the network as a set of t blocks and applies the anticipatory neural network. The first iteration is entered and compared with the desired output, if both are equal; This means that he got the right output. If there are faults (error), adjust the weights of each processing element (node) by applying a neural feedback network for the same block (iteration) (that is, during the training process) these weights are adjusted for accuracy and coverage optimal). After learning both MLP and Bp-NN, the output of the iterations learned from the network points to genetic algorithms GA directly to begin the training phase to choose the best chromosome. GA treats the entries of the Bp-NN and (MLP) networks as a chromosome. Finally, the best chromosome has a good physical shape is the best solution. The MLP network was used for image classification. Bp-NN has been used

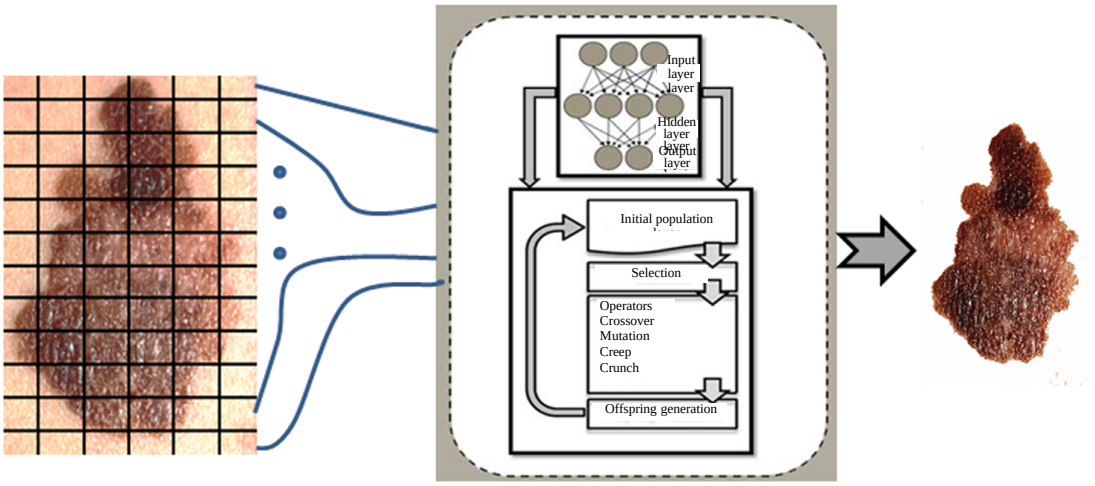

Fig. 4: Sample iteration learned by hybrid proposed system algorithm

for the training of neural networks. According to the evolutionary algorithm, a genetic algorithm begins with a population collection of individuals which evolves towards optimal solutions through genetic operators (selection, crossing, mutation), inspired by biological processes. Each element of the population is called a chromosome and it encodes a point in the search space. The research is guided by a fitness function designed to assess the quality of each individual. The effectiveness of a genetic algorithm is related to the ability to define a good fitness function. The optimization will involve the random search of optimal values of the weights attributed to the connections between neurons in the network where each processing element represents a network of neurons with a particular set of weights. The purpose of the hybrid algorithm is to find the population that produces the smallest value of the error function. In the following table, the infrastructure of the neural network (GA) and the image processing. In the steps below, a complete description of the unsupervised neural networks and the algorithm of the proposed system's genetic algorithm.

RESULTS AND DISCUSSION

Algorithm 1; Genetic algorithm:

The experimental outcomes of the proposed system can be explained as follows:

10 input parameters of melanoma (x_1, x_2, x_3, …, x_n)
Begin Bp-NN Training and Learning
Compare Input parameter with parameter of melanoma image/desired out
If Matching then 30, Else Weight adjusted : Goto step 10
30 Start GA Training
40 GA Operators Appling
Compute the neurons error
If tolerance < > Convergence then goto step 40
Best Solution (Correct Solution)

Table 1: Results acquired by (RBF-NN) training parameters

No. of melanoma disease images	Successful matching	Unsuccessful matching	Training time by second	Image efficiency (%)
6	4	2	0.28	55.00
14	11	3	0.36	80.80
20	17	3	0.39	72.00
22	16	6	0.47	79.55
25	19	6	0.51	81.01

Table 2: Results acquired by (GA) training parameters

No. of melanoma disease images	Successful matching	Unsuccessful matching	Training time by second	Image efficiency (%)
6	5	1	0.24	59.00
14	11	3	0.32	71.05
20	14	6	0.47	72.20
22	16	6	0.51	83.10
25	19	6	0.52	82.05

Table 1-3 of proposed model are organized as follows; Table 1 and 2, experimental results of training parameters acquired by RBF-NN and GA each individually. Table 3, show experimental result by applying NN-GA System.

The increase in the number of training iterations leads to a consumption of effort and time. On the other hand, selecting a limited number of iteration workouts could lead to training outside the network and therefore, to their inability to discriminate. Increasing the amount of training in neural networks leads to reduce the Error Rate (ER) to the desired output. If the network has not reached the desired result, the process of weight adjustment begins at this point in each iteration. The process of weight adjustment includes the propagation of the advance and the return to the best values of the output, so, that the weight adjustment includes all the nodes in the neural network. Through training, the network continuously scales the adjustment associated with each node to produce an output closer to the desired output and this is

Table 3: Experimental result by nn_ga (Hybrid model) system

No. of melanoma images	Hybrid model matching	Error rate	Time training (sec)	Mean accuracy (%)	RMSE
6	5	2	0.10	81	0.30
14	13	2	0.31	79	0.23
20	18	3	0.21	82	0.45
22	20	4	0.40	81	0.35
25	23	6	0.50	95	0.46

based on the Learning Rate (LR). One of the main reasons for non-neuronal network learning is not choosing the right LR ratio. The LR is used to control the rate of increase or decrease in the value of weight values in the learning phase. The training step continues with all the node entries and the weight adjustment related to each node in the network. The output value is closer to the desired output. After reviewing the results by training neural networks, genetic algorithms and the hybrid system, keep in mind that the training results for hybrid systems are better than those of other systems due to the hybrid properties of the neural network. and genetic algorithms. artificial intelligence systems and demonstrated it through the previous results.

By comparing the results with previous experience in the same field using GA to determine the best number of neural network learning sets which have the best weighting aggregates which makes the network more effective by hybridization.

CONCLUSION

In this research, a system developed for prognosis melanoma disease of a patient. RBF-NN and GA have been applied in different areas of medical research due to provide more results that are accurate when it's used in medical tests and diagnosis. The prognosis is done based on historical melanoma database. The RBF-NN and GAs are capable of learning through examples and to generalize with the power of pattern recognition and distinction tasks. They are mathematical models used for understanding and prognosis complex and chaotic dynamics in complex biological systems. Development of a system by using ANN technology and GA for the prognosis of melanoma with high accuracy. To achieve high accuracy, the system builds by the mixed genetic algorithm will be involved with neural network technology.

REFERENCES

Amarathunga, A., E.E. Ala, G.A. Beysekara and C. Amalraj, 2015. Expert system for diagnosis of skin diseases. Intl. J. Sci. Technol. Res., 4: 174-178.

Bakpo, F.S. and L.G. Kabari, 2011. Diagnosing Skin Diseases Using an Artificial Neural Network. In: Artificial Neural Networks-Methodological Advances and Biomedical Applications, Suzuki, K. (Ed.). InTech, London, UK. isBN:978-953-307-243-2, pp: 254-270.

Barati, E., M. Saraee, A. Mohammadi, N. Adibi and M.R. Ahmadzadeh, 2011. A survey on utilization of data mining approaches for dermatological (Skin) diseases prediction. J. Sel. Areas Health Inf., 2: 1-11.

Bullinaria, J.A., 2004. Radial basis function networks: Introduction. Neural Comput. Lect., 13: L13-2-L13-16.

Goldberg, D.E., 2002. Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesle, Boston, Massachusetts, USA.,.

Gound, R.S., S.G. Priyanka, B.G. Jyoti and P.K. Wagh, 2018. Skin disease diagnosis system using image processing and data mining. Intl. J. Comput. Appl., 179: 38-40.

Jamison, D.T., J.G. Breman, A.R. Measham, G. Alleyne and M. Claeson *et al*., 2006. Disease Control Priorities in Developing Countries. 2nd Edn., The World Bank, Washington, D.C., USA. isBN:9780821361795, Pages: 1404.

Kadhim, Q.K., 2017. Classification of human skin diseases using data mining. Intl. J. Adv. Eng. Res. Sci., 4: 159-163.

Kaur, S. and D.R. Bawa, 2015. Future trends of data mining in predicting the various diseases in medical healthcare system. Intl. J. Energy Inf. Commun., 6: 17-34.

Kumar, D.S., G. Sathyadevi and S. Sivanesh, 2011. Decision support system for medical diagnosis using data mining. Intl. J. Comput. Sci. Issues, 8: 148-153.

Michel, F. and L.K. Charles, 1999. Industrial Applications of Genetic Algorithms. CRC Press, Boca Raton, Florida isBN:0-8493-9801-0, Pages: 339.

Munnoli, P.S. and A.U. Bapat, 2013. Clustering algorithms for radial basis function neural network. J. Trans. Electr. Electron. Eng., 1: 113-116.

Parvin1, S.R. and O.A.M. Jafar, 2017. Prediction of skin diseases using data mining techniques. Intl. J. Adv. Res. Comput. Commun. Eng., 6: 313-318.

Patel, S. and H. Patel, 2016. Survey of data mining techniques used in healthcare domain. Intl. J. Inf., 6: 53-60.

Priyanga, A. and S. Prakasam, 2013. The Role of Data Mining-Based Cancer Prediction System (DMBCPS) in cancer awareness. Intl. J. Comput. Sci. Eng. Commun., 1: 54-61.

Sudha, J., M. Aramudhan and S. Kannan, 2017. Development of a mathematical model for skin disease prediction using response surface methodology. Biomed. Res., 2017: S355-S359.

WHO., 2011. Global Status Report on Noncommunicable Diseases 2010. World Health Organization, Geneva, Switzerland is BN-13: 9789241564229, Pages: 162.