Azerbaijan Journal of High Performance Computing, Vol 5, Issue 1, 2022, pp. 112-130
https://doi.org/10.32010/26166127.2022.5.1.112.130

Application of Deep Boltzmann Machine in Diagnosis Processes of Hepatitis Types B & C

Hadis Oftadeh[1] and Mohammad Manthouri[2]

[1] *Islamic Azad University, Tehran, Iran, oftadeh@azad.ac.ir*
[2] *Shahed University, Tehran, Iran, mmanthouri@shahed.ac.ir*

*Correspondence:
Hadis Oftadeh, Islamic Azad University, Tehran, Iran, oftadeh@azad.ac.ir

Abstract

Correct diagnosis of diseases is the main problem in medicine. Artificial intelligence and learning methods have been developed to solve problems in many fields, such as biology and medical sciences. Correct diagnosis before treatment is the most challenging and the first step in achieving proper cures. The primary objective of this paper is to introduce an intelligent system that develops beyond the deep neural network. It can diagnose and distinguish between Hepatitis types B and C by using a set of general tests for liver health. The deep network used in this research is the Deep Boltzmann Machine (DBM). Learning components within Restricted Boltzmann Machine (RBM) lead to intended results. The RBMs extract features to be used in an efficient classification process. An RBM is robust computing and well-suited to extract high-level features and diagnose hepatitis B and C. The method was tested on general items in laboratory tests that check the liver's health. The DBM could predict hepatitis type B and C with an accuracy between 90.1% and 92.04%. Predictive accuracy was obtained with10-fold cross-validation. Compared with other methods, simulation results on DBM architecture reveal the proposed method's efficiency in diagnosing Hepatitis B and C. What made this approach successful was a deep learning network in addition to discovering communication and extracting knowledge from the data.

Keyword: Deep learning, Restricted Boltzmann Machine, Hepatitis, Neural Network, Classification.

1. Introduction

The most critical gland of the human body is the liver, a vital organ due to its functionality, and hepatitis is the inflammation of the liver. The hepatitis viruses represent a significant public health problem worldwide and replicate primarily in the liver, causing related diseases with a broad range of clinical manifestations (Quer, J., Rodríguez-Frias, F., et al., 2017). According to the world health statistics in 2015, the seventh cause of 1.34 million deaths was hepatitis (World Health Organization, 2017). Despite the importance of this disease, it neglects as a health factor. Hepatitis B and

C viruses in chronic conditions can lead to other liver diseases, including cirrhosis and liver cancer-chronic (Baumert, T. F., Berg, T., Lim, J. K., & Nelson, D. R., 2019). A blood test would be helpful to achieve a correct diagnosis and check if specific antiviral antibodies exist (Razavi, H., 2020). The accurate diagnosis could be the impact factor in optimal treatments, so fewer diagnosis errors can support patients to receive appropriate care.

Nowadays, medical society prefers to save time in the interpretation and communication of clinical decision. The presence of new technologies like medical data analyses has been abundant, and intelligent algorithms produce significant results. A Multilayer network can train a neural network to diagnose diabetes (Gadekallu, T. R., Khare, N., et al., 2020), chronic obstructive pulmonary disease (Nunavath, V., Goodwin, M., Fidje, J. T., & Moe, C. E., 2018, September), and Hepatitis (AbuSharekh, E. K., & Abu-Naser, S. S., 2018). Learning methods can extract features from ultrasound images to classify and diagnose the staging of chronic liver diseases (Ahn, J. C., Connell, A., Simonetto, D. A., Hughes, C., & Shah, V. H., 2021) or for adjudging the degree of liver fibrosis in chronic hepatitis C (Durot, I., Akhbardeh, A., Sagreiya, H., Loening, A. M., & Rubin, D. L., 2020), also liver segmentation from CT volumes (Liao, M., Zhao, Y. Q., et al., 2017) and MRI images (Bereciartua, A., Picon, A., Galdran, A., & Iriondo, P., 2016).

Machine Learning algorithms can learn from data and make considerations and predictions based on a data set. It is usual to use the advantages of methods inside each other such as benign or malignant classification of breast cancer tumors by Genetically Optimized Neural Network (Bhardwaj, A., & Tiwari, A., 2015). Also, extraction of morphologic features is utilized as the input of a multilayer feed-forward neural network (Movahedi, M. M., Zamani, A., et al., 2020), Elsewhere; there could be a combination of many techniques such as a support vector machine and simulated annealing used in both methods for detection of some hepatitis (Akbar, W., Wu, W. P., et al., 2020). Experiences have shown that combination methods bring data problems to be covered by another part. Nevertheless, it still might be better to use data mining methods before data provision as the inputs of a learning algorithm. Some medical equipment is based on intelligence algorithms in the diagnosis domain like Elastography and application in the diagnosis of hepatic fibrosis stage (Chen, Y., Luo, Y., et al., 2017). However, Elastography extracts the features whose relationship was discovered so far but is not sufficiently accurate. In this regard, deep learning is an admirable ability to gain powerful features with linear and nonlinear relationships. The conventional method used in training a restricted Boltzmann machine is the contrastive divergence (Upadhya, V., & Sastry, P. S., 2019). This technique has an impact on training performance. This advantage is that a gradient estimate is definite, and a more significant training rate can be used (Li, F., Gao, X., & Wan, K., 2018). In addition, Principle Component Analysis (PCA) as a reduction method can be used to classify hepatitis (Nilashi, M., Ahmadi, H., Shahmoradi, L., Ibrahim, O., & Akbari, E., 2019); PCA extracts linear features, apparently, unable to take nonlinear ones; contrarily, RBM can extract them (Bu, Y., Zhao, G., Luo, A. L., Pan, J., & Chen, Y., 2015). Auto-encoder, as

one of the Deep learning methods in dimension reduction, is much more favorable than previous methods for dimension reduction, including principal component analysis (Xiao, Y., Xing, C., Zhang, T., & Zhao, Z., 2019). Though an auto-encoder can be very desirable to extract features using deep learning, this should take into consideration that good results require more time to optimize the model.

Solving many issues is simple for computers. Deep learning seeks to learn experiences and understands the world through a series of concepts (Wani, M. A., Bhat, F. A., Afzal, S., & Khan, A. I., 2020). AI is working on new learning methods to facilitate diagnosis processes with reliable accuracy. In recent years, deep learning, a branch of Machine Learning, has successfully been working due to excellent ability in learning, which is the main feature of these methods.

There are five types of hepatitis viruses: A, B, C, D, and E; among them, types A, B, and C are more common. The features and symptoms of hepatitis B and C are so similar that choosing them to use in diagnoses is of great importance. A Deep Neural Network is expected to help achieve this goal.

The RBM creates an independent framework for competing in nonlinear classifications; Therefore, a discriminative RBM can also be prosperous of a semi-supervised learning and follow-up study of a satisfactory analysis (Varsamou, M., & Antonakopoulos, T., 2019, September). It can argue that this method is used as a method of nonlinear independent classifications that would be very useful in medical areas (Mahmoud, A. M., Alrowais, F., & Karamti, H., 2020). Learning methods that are flexible, able to extract attributes automatically, and obtain nonlinear dependencies would be essential in the field of medicine (Wang, C., Tan, X. P., Tor, S. B., & Lim, C. S., 2020).

This research studies the effectiveness of deep learning for classifying hepatitis B and hepatitis C viruses. This method is efficient and fast to gain desirable results. Additional tests are needed to diagnose hepatitis. However, this research proves that deep learning can classify hepatitis B and hepatitis C by having a set of general laboratory tests for liver health and gaining satisfactory accuracy. The purpose of deep learning is to solve some problems understandable to man's mind but cannot be described by computer models. This article shows that the results of the medical test can be correctly modeled in deep learning, so some of the additional laboratory tests for the diagnosis of hepatitis may remove in some cases. Also, compared with other methods that use feature engineering for classification, DBM is more reliable because of its extreme ability to extract nonlinear features.

This paper is organized as follows. Section 2 introduces RBM as a training algorithm in DBM architecture. Database description is provided in Section 3. In Section 4, the quality of the presented passage is examined, and the significance of the results of the work investigates the effect of DBM on classifying hepatitis B and hepatitis C. The significance of the results is devoted in Section 5, and the conclusion of this work appears in Section 6.

2. The Structure of RBM

Deep learning should be able to model human knowledge. In this part, RBM is

introduced as the fundamental algorithm. The main idea is to declare training of components of the Deep Neural Network and its architecture.

2.1. Restricted Boltzmann Machine

The Boltzmann machine (BM) is a neural network with one hidden layer. BM consists of full connectivity between neurons in the same layer and between two respective layers. Figure 1-a shows connections in BM.

There is a type of BM called Restricted Boltzmann Machine (RBM) in case the connectivity of the same layer's neurons is ignored or limited. The learning process in the RBM is faster than in the general Boltzmann Machine. RBM is shown in Fig. 1-b.

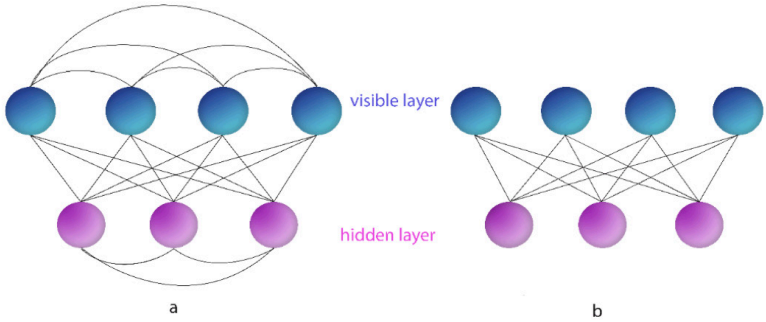

Fig. 1. Boltzmann machine vs. Restricted Boltzmann machine

BM and RBM reconstruct inputs in the output layer. Another expression of data would appear in the hidden neurons at the end of training the BM. In other words, the invisible layer consists of feature extraction that the input would substitute. The BM is a probabilistic model based on energy as follows:

$$p(\theta) = e^{-E(\theta)} \Big/ \sum e^{-E(\theta)} \tag{1}$$

Where θ is the parameter, E is the energy function, and p shows the probability. The progress of E is inversely proportional to P. Hence, a decrease in energy leads to an increase in probability. Eventually, the Boltzmann machine would be detected the dependency of the variable. The DBM is a methodology that uses RBM in which the hidden layers organize sequentially and deeply. The RBM is a Markov Random Fields and is an undirected graphical model that follows Markov's property. According to Markov's attribute, every two nodes are conditionally independent of all other neighbors. The energy function of the RBM is expressed as:

$$E(v,h) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{i,j} h_i v_j - \sum_{j=1}^{m} b_j v_j - \sum_{i=1}^{n} c_i h_i \tag{2}$$

Where v expresses visible neurons $v_i \in \{0,1\}$, h definite hidden neurons, $h_i \in \{0,1\}$,

BM and RBM reconstruct inputs in the output layer. Another expression of data would appear in the hidden neurons at the end of training the BM. In other words, the invisible layer consists of feature extraction that the input would substitute. The BM is a probabilistic model based on energy as follows:

$$p(\theta) = e^{-E(\theta)} \Big/ \sum e^{-E(\theta)} \tag{1}$$

Where θ is the parameter, E is the energy function, and p shows the probability. The progress of E is inversely proportional to P. Hence, a decrease in energy leads to an increase in probability. Eventually, the Boltzmann machine would be detected the dependency of the variable. The DBM is a methodology that uses RBM in which the hidden layers organize sequentially and deeply. The RBM is a Markov Random Fields and is an undirected graphical model that follows Markov's property. According to Markov's attribute, every two nodes are conditionally independent of all other neighbors. The energy function of the RBM is expressed as:

$$E(v,h) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{i,j} h_i v_j - \sum_{j=1}^{m} b_j v_j - \sum_{i=1}^{n} c_i h_i \tag{2}$$

Where v expresses visible neurons $v_i \in \{0,1\}$, h definite hidden neurons, $h_i \in \{0,1\}$, $w_{i,j} \in \tilde{}$ is weights that connect visible unit i and hidden unit j, c_i and b_j are the bias term of the visible unit i and hidden unit j, n and m are the numbers of hidden units, and visible units.

The probability distribution is acquired using equation 3:

$$p(v,h) = \frac{1}{z} e^{-E(v,h)} \tag{3}$$

In which z is the partition function. According to equations 1 and 3, the partition function would be as follows:

$$Z = \sum_{x} e^{-E(x)} \tag{4}$$

The Restricted Boltzmann machine is based on conditional probability. In this way, hidden and visible neurons are conditionally independent so that they would be described as:

$$p(h \mid v) = \prod_{i=1}^{n} p(h_i \mid v) \tag{5}$$

$$p(v \mid h) = \prod_{j=1}^{m} p(v_j \mid h)$$

The target is the activation of each neuron. So, the probability would be as follow:

$$p(h_j = 1 \mid v) = \sigma(b_j + \sum_i v_i w_{ij}) \tag{6}$$

$$p(v \mid h) = \prod_{j=1}^{m} p(v_j \mid h)$$

The target is the activation of each neuron. So, the probability would be as follow:

$$p(h_j = 1 \mid v) = \sigma(b_j + \sum_i v_i w_{ij}) \quad (6)$$
$$p(v_i = 1 \mid h) = \sigma(a_i + \sum_j h_j w_{ij})$$

Where is the sigmoid activation function?

It is necessary to determine how the parameters of the networks are valued to diagnose hepatitis. Likewise, what criteria would be advisable to be used in assessments for the superiority of θ parameters? To answer this question, we usually use statistical models. P and θ as parameters describe a probability distribution of P (θ); frequently, we need to amplify and maximize $p(v;\theta)$. v is the network's input, and probability exponential is logistic with parameters $w^T h + b$.

Backpropagation is a somewhat more professional approach to doing gradient descent, a method that involves partial derivatives by dynamic programming.

2.2. Training an RBM

An RBM would be trained by approximate log-likelihood by defining as:

$$L(\theta) = \sum \log p(v \mid \theta) \quad (7)$$

The goal is to maximize the log-likelihood of training an RBM, so the gradient alongside the model parameters needs to be computed (Taherkhani, A., Cosma, G., & McGinnity, T. M., 2018):

$$\frac{\partial \ln \ell(\theta \mid v)}{\partial \theta} = -\sum_h p(h \mid v) \frac{\partial E(v,h)}{\partial \theta} + \sum_{v,h} p(v,h) \frac{\partial E(v,h)}{\partial \theta} \quad (8)$$

The calculation of the first expression in equation (8) is easy because it can be simplified by factoring as follows:

$$\sum_h p(h \mid v) \frac{\partial E(v,h)}{\partial wij} = \sum_h p(h \mid v) hivj = \sum_h \prod_{k=1}^{n} p(h_k \mid v) p(h_{-i} \mid v) h_i v_j$$
$$= \sum p(h_i \mid v) h_i v_j \underbrace{\sum_{h_{-i}} p(h_{-i} \mid v)}_{=1} = p(H_i = 1 \mid v) v_j = sig(\sum_{j=1}^{m} w_{i,j} v_j + c_i) v_j \quad (9)$$

And the second part of the equation (8) means the expected energy gradient under the Boltzmann distribution, which can be written as follows

$$\sum_v p(v) \sum_h p(h \mid v) \frac{\partial E(v,h)}{\partial \theta}$$

or

$$\sum_{h} p(h) \sum_{v} p(v \mid h) \frac{\partial E(v,h)}{\partial \theta} \tag{10}$$

Contrastive Divergence (CD) (Zhang, Y., Li, P., & Wang, X., 2019) and Persist Contrastive Divergence (PCD) (Li, F., Gao, X., & Wan, K., 2018), which are based on CD, can be used to get the desired approximation of the gradient from the model distribution. The typical way is, using a Markov Chain. After running the chain, only a few steps are enough to train the model and estimate to be made.

K-fold Cross-Validation

Cross-validation is a strategy to assess predictive models by dividing the original sample into the training and test set. This technique repeats the experiment multiple times, using all the different parts of the training set as validation sets.

In k-fold cross-validation, the samples are divided into k-size subsamples. Of the k subsamples, a single subsample is held as the validation data for testing the model, and the others are used as training data.

2.4. Mini Batches

Data is divided into mini-batches to train an RBM. Since data in one epoch is large to feed the deep network at once, we divide it into several smaller batches. The advantage of using mini-batches is the primary idea in stochastic gradient descent (Bottou, L., 2010). The use of mini-batches can cause to injection of the appropriate noise in each update of stochastic gradient descent.

2.5. Deep Boltzmann machine (DBM)

As mentioned, an RBM is a probabilistic model with a hidden layer for modeling the distribution of visual neurons. The accumulation of RBMs creates deep belief networks (Taherkhani, A., Cosma, G., & McGinnity, T. M., 2018) for hierarchical processing. Most of the changes and modifications made to improve this type of network lead to the optimal training of the RBM. A Deep Boltzmann machine is a graphical model with undirected connections between the variables and includes the accumulation of several RBMs. It is a specific bipartite diagram in which the vertices relating to each layer are associated with the layer instantly above and immediately beneath it.

DBM joint distribution is given by the exponential expression as follows:

$$P(\upsilon, h^{(1)}, ..., h^{(d)}) = \frac{1}{Z(\theta)} \exp(E(\upsilon, h^{(1)}, ..., h^{(h)})) \tag{11}$$

E is energy and given by:

$$E(\upsilon, h^{(1)}, ..., h^{(d)}) = \sum h^{(i)^T} W^{(i+1)} h^{(i+1)} + \upsilon^T W^{(1)} h^{(1)} \tag{12}$$

DBMs have the strength of training internal representations that acquire increasingly complex (Taherkhani, A., Cosma, G., & McGinnity, T. M., 2018).

The training of the model is carried out in three steps.

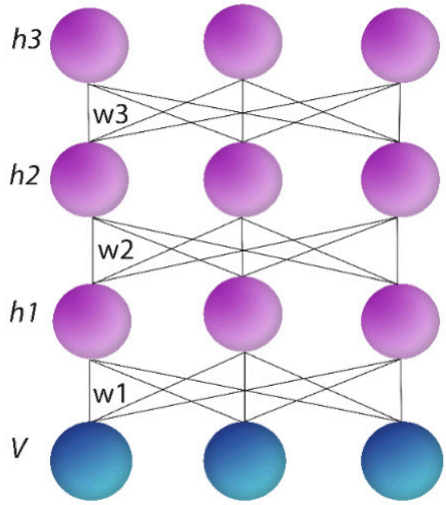

Fig. 2. Three-layer Deep Boltzmann Machine

Pre-training: each layer in DBM is an RBM; they should be trained separately and respectively. The output of each layer is the following RBM input. At first, variables initialize randomly. The RBM map the features into hidden layer neurons. Training RBM takes place in the positive, feed-forward, and negative phases to reconstruct the data. By the end of the training, the RBM values of the hidden layer are intended purpose output with high-level features.

Unrolling: in this step, the entire layers are considered seamlessly. The input layer to the output layer calculates feed-forward with the weights obtained during the pre-training phase. Classification would be in the final layer.

Fine-tune: the last step, called fine-tunes, uses gradient descent in training and supervised classification.

3. Data Structure

PKDD02 is the data collection of instances of hepatitis B versus hepatitis C PKDD02 is a database described as a discovery challenge in 2002 and 2003 of PKDD (Practice of Knowledge Discovery in Databases) (ECML/PKDD, 2002). The values of the original dataset are absolute and have many null values. Because of clean and accessible, a modified version of PKDD02 (Dataset, 2022) is used in this article.

It contains general laboratory examinations of liver health and defines hepatitis B and hepatitis C-infected patients. In addition to the laboratory testing, necessary parameters such as age and sex are considered.

The values in this database are integer numbers and seem to describe a level (or class) of laboratory interpretation. Data would be usable in the DBM as input converted to binary data. 206 samples of hepatitis B and 484 samples of hepatitis C exist in the database, and concerning the periodic test, we access more samples to be the input

of the deep network. The total record is 5370, of which 3761 are for hepatitis C, and 1609 for hepatitis B. Separating data into training and testing sets are shown in Table 1.

Table 1. Data dividing for training and testing

All Data	Training data	Test Data
5370	4680	690

This research data is divided into mini-batches to gain the benefits and reduce the gradient's variance. As shown in Table 2, if the size of each batch is 30, then 156 batches are allocated to training data, and 23 batches are for test data.

Table 2. Divide testing and training data into small packages

Total batches	Size of each batch	Training batch number	Test batch number
179	30	156	23

This research data is divided into mini-batches to gain the benefits and reduce the gradient's variance. As shown in Table 2, if the size of each batch is 30, then 156 batches are allocated to training data, and 23 batches are for test data.

The data in this research is an integer which represents a range. For example, if age considered being the baby to a teenager would be 1, young to an adult would be 2, and the elder would be 3. This limitation has reduced the dispersion of numbers. The relationship between the various attribute is evident. Since the base algorithm in this research is RBM. In this issue, 12 main features are onside for data, which includes the following information:

Albumin: Albumin is the main protein in the blood and is a family of globular proteins which is water soluble. In the human body, it is one of the essential proteins in the plasma.

GOT enzyme: An enzyme found mainly in the heart muscle and liver with moderate amounts in the skeletal muscle, kidneys, and pancreas.

GPT: An enzyme found primarily in liver cells and helpful in detecting liver cellular degeneration.

Bilirubin: One of the biliary yellow pigments that resulted from the breakdown and expected hemoglobin degradation.

Cholinesterase enzyme is an enzyme that helps the proper functioning of the nervous system.

Total protein: The whole protein in the blood.

Total cholesterol: Cholesterol is a fat-free substance found in all cells in the body

There is also the thymol turbidity and opacity of zinc sulfate, both liver function tests; in addition to laboratory parameters, gender and age be discussed.

4. Experiments

In this research, the Deep network is trained with two hidden layers. The number of neurons required should evaluate in error. However, 300 neurons for the first layer and

450 neurons for the second layer are considered, which would change because of the error rate.

RBM is a probabilistic model with a hidden layer that models a distribution of visual neurons. The accumulation of RBMs creates deep networks for hierarchical processing. Therefore, improving this type of network leads to the optimal training of the RBM. The training model contains three primary steps:

Pre-training: deep neural network comprises multiple RBM which train individually, and each layer's output is the following RBM input.

Unrolling: in this part, the entire network stage is considered seamlessly, and computing from the input layer to the output layer progressively with the weights obtained in the pre-training phase,

Fine-tune

The initial values of the weight and bias are generated randomly. After training in each epoch, the program's parameters, including weight and bias, are updated. The learning rate is 0.05 in this study. It is worth mentioning that the learning rate has already been evaluated to be 0.1, 0.05, and 0.01 and that the minimum error has been 0.05. The learning rate is constant only in the first layer, while in the second layer and during the training of the RBM, the learning rate is a coefficient of the final learning rate in the previous epoch. This coefficient is assumed to earn dynamically per epoch. The minimum and maximum values for the momentum are 0.5 and 0.9, respectively; the former is allotted to the initial epochs, and the latter is used for the rest. Also, Weight decay is constant at a value of 0.01.

Features obtained from training DBM should be classified to diagnose hepatitis B, and C. Classification performs better on cross-validated data when miss-classification is taken as an evaluation parameter. In this research, 10-fold cross-validation is used. It shows 68 miss-classification in 10 execution. We determined the miss-classification of a classifier and MSE for training each layer of DBM and, first, computing the total error (miss-classification rate or MSE) of 10-fold cross-validation and second, computing the average error per fold by entailing the total error divided by the number of folds.

Figure 3-a shows the reconstruction of a data category in the first RBM after 10 epochs, and figure 3-b shows the reconstruction of the same data after 100 periods.

As depicted in figure 3-b, finally, after training on 100 epochs, data reconstructed the map on the original data. The vertical axis in Figure3 shows the number of data in one batch, and the horizontal axis represents the number of features.

RBM can learn another form of data and model the distribution of observable layers on hidden layers. A meticulous reconstruction leads to the algorithm's smooth and predictable performance across the instance space.

The Mean Square Error (MSE) to reconstruct data in RBM1 and RBM2 by five iterations describes in table3.

After completing training RBM1 by 100 epochs and 150 iterations in RBM2, MSE would be as follow-on table 4 and also shown in Figure 4:

Fig4 represents the MSE in 150 iterations per batch in both layers. As shown in

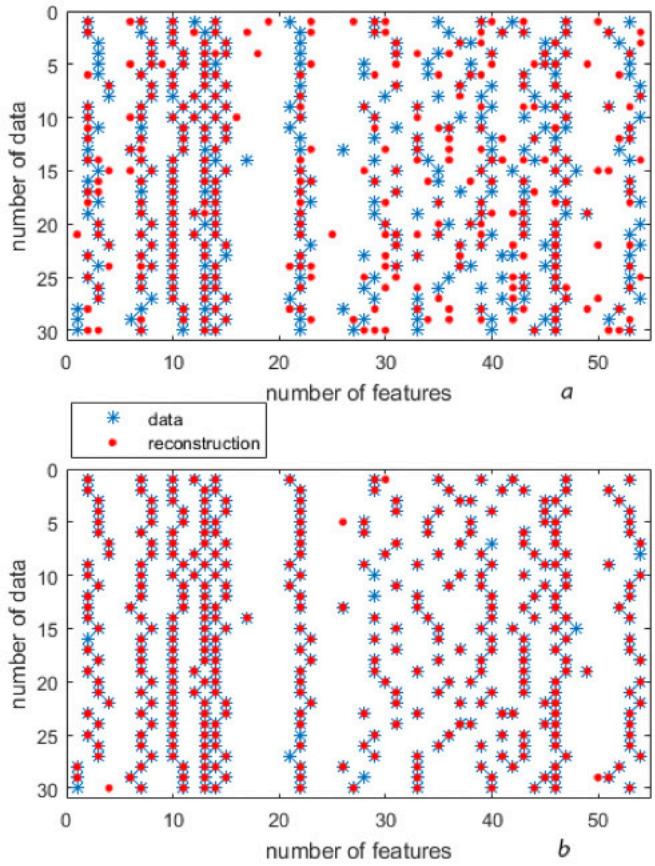

Fig. 3. data in a batch versus reconstruction data in the same batch

Figure 3-a shows the reconstruction of a data category in the first RBM after 10 epochs, and figure 3-b shows the reconstruction of the same data after 100 periods.

As depicted in figure 3-b, finally, after training on 100 epochs, data reconstructed the map on the original data. The vertical axis in Figure3 shows the number of data in one batch, and the horizontal axis represents the number of features.

RBM can learn another form of data and model the distribution of observable layers on hidden layers. A meticulous reconstruction leads to the algorithm's smooth and predictable performance across the instance space.

The Mean Square Error (MSE) to reconstruct data in RBM1 and RBM2 by five iterations describes in table3.

Table 3.MSE in 10 execute after 5 repetitions

RBM	MSE	MSE in each batch
Layer 1	28.6	0.19
Layer 2	5.1	0.04

After completing training RBM1 by 100 epochs and 150 iterations in RBM2, MSE would be as follow-on table 4 and also shown in Figure 4:

Table 4. The MSE for 10 times run after 100 repetitions in the first layer and 150 repetitions in the second layer

RBM	MSE	MSE in each batch
RBM1	19.47	0.12
RBM2	4.5	0.03

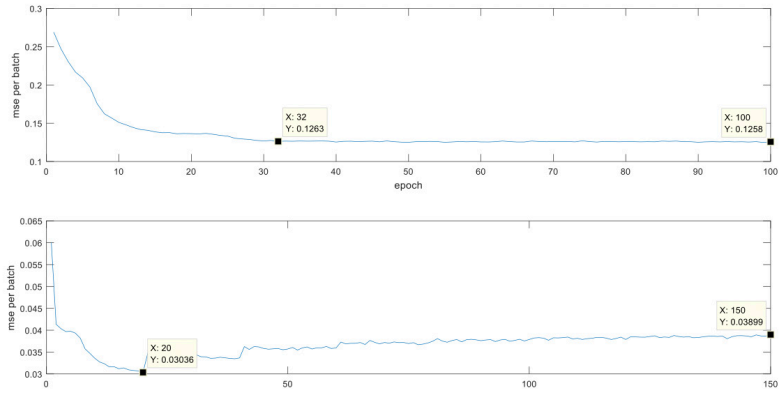

Fig. 4.The mean square error in the first and second layer of the Restricted Boltzmann Machine

Fig4 represents the MSE in 150 iterations per batch in both layers. As shown in figure 4, MSE in RBM2 after 20 epochs has begun to increase; however, the increase is at 0.008. Also, RBM1 after 30 iterations did not significantly increase or decrease. The error rate may increase or decrease due to the number of hidden layers or learning rates.

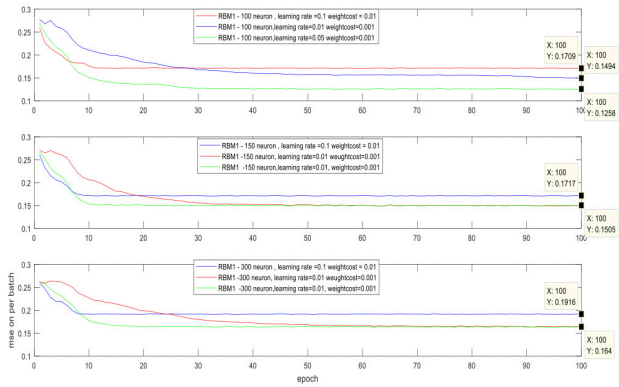

Fig. 5.Investigating the effect of the number of neurons in the hidden layer and learning rates on the MSE in the first RBM

Fig 5 declares the effect of the number of neurons in the hidden layer and learning rates on the MSE per batch in each epoch in the first RBM. To assess the performance of configurators in the principled training, we can estimate the error rate for the reconstruction of RBM based on hidden layer neurons and learning rate. Additionally, we repeated the experiment 10 times for neurons, learning rate, and weight cost, the effect of the number of hidden neurons with the different learning rates in RBM1 is illustrated in Figure 5; the lowest error rate obtained with 100 neurons in the hidden layer and a training rate of 0.05. We use configurators to bias training data to high-performance regions.

Fig 6 defines the MSE rate in 10 executed per batch in each epoch in the second RBM. The MSE rate for the second layer in Figure 6 shows that convergence in an RBM is fast, and there is no significant increase or decrease in error rates after epoch 20.

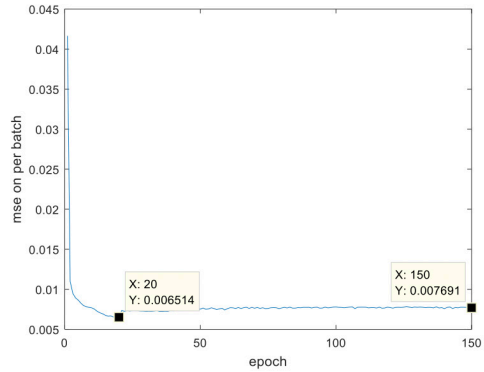

Fig. 6. The MSE rate after 10 execute in the second RBM

A comparison of the effect of using PCD and CD on the MSE in the first RBM is shown in Figure 7. MSE in using PCD learning as the training algorithm for the first RBM would be less than the results of the CD.

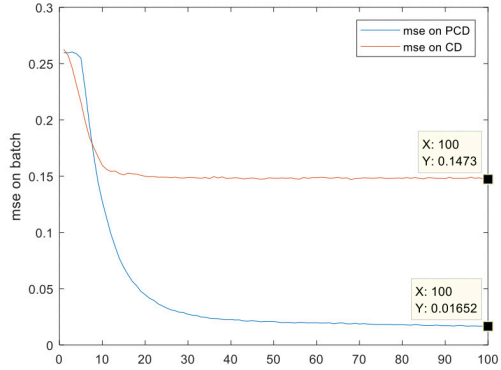

Fig. 7.Comparison of the effect of using PCD and CD on the MSE in the first RBM

We optimized for misclassification rate on 10-fold cross-validation on data. We adopt the evaluation of model classification accuracy in Figure 8, which shows the error rate on the test data

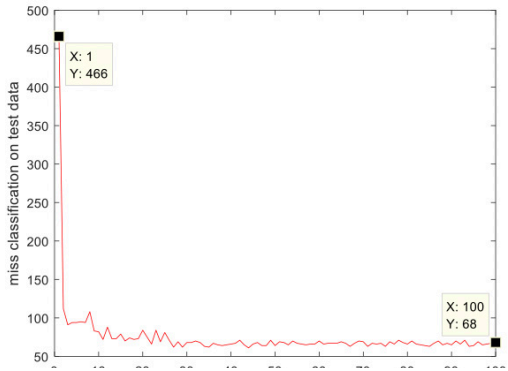

Fig. 8. The classification error rate on the test data

The use of RBM facilitates classification and minimizes errors. If CD trains both layers, the classification error in test data would be 90.1; otherwise, if PCD trains RBM2, the classification error would be 92.04.

Deep Boltzmann Machine gains better scores in classification than other learning machine methods. Results of some methods developed on PKDD02 are depicted in Table 6.

Table6.DBM performance compared to other models

Input Data	Algorithm	Accuracy
PKDD02	SVM	73.48
	Nonlinear SVM (RBF kernel)	71.08
	DBM-CD	90.1
	DBM-PCD	92.04
	Aleph (Ferreira, C. A., Gama, J., & Costa, V. S., 2011, October)	78
	DARA (Sia, F., Alfred, R., & Chin, K. O., 2013, August)	81.4
	E-NB (Schulte, O., Bina, B., et al., 2013, April)	78
	HNBC (Schulte, O., Bina, B., et al., 2013, April)	78
	PIC (Schulte, O., Bina, B., et al., 2013, April)	80

5. Discussion

Diagnosis of hepatitis virus from the general liver blood test is not easy. Hence, we have to extract the data's features that lead to a suitable model for diagnosing hepatitis B and C with the desired accuracy. The high ability of DBM to generate new features from limited series of features has been the main reason for choosing DBM. Since deep learning reduces the need for feature engineering, which is time-consuming learning, deep learning can create features without human intervention. Also, DBM is more efficient for problems because the mathematical formulation is impossible.

Relevant features that accurately predict algorithm performance in a DNN are usable and reliable. Based on the results, the propounded deep learning model yield satisfactory results to predict hepatitis B or C by general laboratory test up to 92.04accuracies. Hepatitis can cause both acute and chronic disease, accurate and fast diagnosis is vital. It is highly desirable to develop a model that allows dramatic performance in experimentation, cutting down the time required for diagnosing hepatitis without needing specific tests. The proposed model results confirm that deep learning training discovers unknown and complex relationships between data and fits them to the best class label. Deep Learning methods are favorable to implement because of their high computational capabilities and less required time for classification. They incur high computational costs (Xiao, Y., Wu, J., Lin, Z., & Zhao, X., 2018). Accurate diagnosis of hepatitis is acquirable through technical laboratory tests. However, They are time-consuming and cost a lot.

Medical Data fields are notable because they deal with people's health in a community. One of the challenges in practical medical research is limited access to a large number of patients, so sufficient data is not collected in this regard. If you don't have permission, it is impossible to access data. Points to the importance of deep learning since medical data are usually heterogeneous (Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T., 2018)—deep learning predictive analytics model.

Deep Learning is a technique with enormous amounts of computational power and is increasingly gaining popularity in the fields of medicine. This event is because fast and accurate diagnostics would be gained through deep learning.

The databases in medical fields are relatively small. It is no good reason to avoid the high capabilities of deep learning in small datasets. Over-fitting becomes an even more significant issue in deep learning. If we have a limited amount of data, we risk over-fitting. So it is vital to make these studies more accurate. One strategy for battling over-fitting is regularization. One frequently used technique is weight-decay regularization by adding a term that penalizes large weights. In this paper, training DBM uses the potential benefit of weight decay regularization as a weight cost parameter that should be chosen carefully. The weight decay has no desirable effect if it is too small and too large, so the deep network did not model the data distribution well. The effect of weight cost in training RBM is shown in figure3. Also, employing cross-validation reduces the risk of over-fitting on the test set. To discover how generalizable the model is, we used 10-fold cross-validation. During the training and evaluation of the network for hepatitis B and C diagnosis, the average free energy was about the same on the test and training set, so the probability of over-fitting decreased.

Furthermore, the selection of suitable batch sizes provides some limit on over-fitting. However, so small batches can lead to over-fitting.

The Hepatitis virus has been considered by computational science in many aspects because liver diseases can affect one another. The correct diagnosis of hepatitis B or C described in this article achieves 92.04% accuracy and would be helpful for other researchers, like the diagnosis of liver fibrosis in patients with hepatitis C with 84.4% (Hashem, S., Esmat, G., et al., 2017). The genome sequencing project predicts HBeAg status and DNA level and produces a sufficiently accurate 95% and 94%respectively (Podlaha, O., Revill, P., et al., 2017). Also, the study of the viral genome obtains higher than 90% correctness in classification (Remita, M. A., Halioui, A., et al., 2017).

6. Conclusion

The use of learning methods with the intelligent approach is the basis of artificial intelligence projects. In the learning machine, the problem divides into different computational levels; they evaluate the results of calculations. Feature extraction is possible with many structures and methods that reduce the input data's dimensionality. This research focused on the training on deep learning, and the components of these networks, including the RBM on the classification of hepatitis B and hepatitis C, were examined.

The goal of deep learning is to learn the hierarchy of features. The RBM uses a hidden layer to model data distribution on its observable variables. In other words, RBM maps the data properties into hidden layer neurons. This ability of the RBM allows extracting high-level features from the low-level ones of the data.

A few examples of the RBM's capabilities explain briefly below:

It can code any distribution, while it has computational efficiency.

Because of using the hidden layer, it can also recognize the high correlations.

Because of the limitations of the standard Boltzmann machine, its processing is faster than it is.

It has high computational power.

This research aimed to examine the depth of learning capabilities for modeling human knowledge. We presented a novel approach for constructing a model to diagnose hepatitis B virus and hepatitis C viruses. We developed a comprehensive methodology to enable the quality of medical services. This method generates to help medical society to support patients more than past. This study uses the standard strategy for evaluating deep networks by splitting the data into 10-fold cross-validation. It is used for over-fitting avoidance. A testing set would determine if the network has learned the training patterns and reached the final performance.

Unlike deep learning, other methods have many challenges in selecting or extracting the features. The classification of hepatitis B virus and hepatitis C virus gains the best accuracy in deep learning, reflecting the computational power of deep learning networks.

Deep learning is progressing rapidly and applies to diverse domains of science, such as bioinformatics. Transforming medical biology data into accessible knowledge

is one of the most critical challenges in bioinformatics. High-performance computing in deep learning can be used in analyzing bioinformatics (Min, S., Lee, B., & Yoon, S., 2017) data such as sequences, genes, and protein expression (Larranaga, P., Calvo, B., et al., 2006). Gene Location and structure of the genes can be extracted from genome sequences. Furthermore, gene function and RNA secondary structure prediction are obtained from sequence information. Also, bioinformatics analysis of microarray data is essential for the study of diseases (Yang, Y., Zhong, Z., et al., 2018). It is recommended that genomic sequences be used if someone is interested in hepatitis and deep learning study. However, there are methods for analyzing data, like proteomics and structural prediction of proteins, such as protein-protein interaction (PPI) (Sun, T., Zhou, B., Lai, L., & Pei, J., 2017) in hepatitis viruses. Hepatitis viruses cause disease in humans by changing PPI within the host cells. Prediction PPI can help to comprehend protein function and treatment design. It can be evaluated with particular reference to the case of hepatitis data.

Reference

AbuSharekh, E. K., & Abu-Naser, S. S. (2018). Diagnosis of hepatitis virus using artificial neural network. *International Journal of Academic Pedogogical Research, 2*(11), 1-7.

Ahn, J. C., Connell, A., Simonetto, D. A., Hughes, C., & Shah, V. H. (2021). *Application of artificial intelligence for the diagnosis and treatment of liver diseases. Hepatology, 73*(6), 2546-2563.

Akbar, W., Wu, W. P., et al. (2020). Development of hepatitis disease detection system by exploiting sparsity in linear support vector machine to improve strength of adaboost ensemble model. *Mobile Information Systems,* 2020.

Baumert, T. F., Berg, T., Lim, J. K., & Nelson, D. R. (2019). Status of direct-acting antiviral therapy for hepatitis C virus infection and remaining challenges. *Gastroenterology, 156(2), 431-445.*

Bereciartua, A., Picon, A., Galdran, A., & Iriondo, P. (2016). 3D active surfaces for liver segmentation in multisequence MRI images. *Computer Methods and Programs in Biomedicine, 132,* 149-160.

Bhardwaj, A., & Tiwari, A. (2015). Breast cancer diagnosis using genetically optimized neural network model. *Expert Systems with Applications, 42*(10), 4611-4620.

Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010 (pp. 177-186).* Physica-Verlag HD.

Bu, Y., Zhao, G., Luo, A. L., Pan, J., & Chen, Y. (2015). Restricted Boltzmann machine: *a non-linear substitute for PCA in spectral processing. Astronomy & Astrophysics, 576, A96.*

Chen, Y., Luo, Y., et al. (2017). Machine-learning-based classification of real-time tissue elastography for hepatic fibrosis in patients with chronic hepatitis B. *Computers in biology and medicine, 89,* 18-23.

Dataset (2022). *https://relational.fit.cvut.cz/dataset/Hepatitis.*

Durot, I., Akhbardeh, A., Sagreiya, H., Loening, A. M., & Rubin, D. L. (2020). A new multimodel machine learning framework to improve hepatic fibrosis grading using ultrasound elastography systems from different vendors. *Ultrasound in medicine & biology, 46*(1), 26-33.

ECML/PKDD (2002). https://www.cs.helsinki.fi/events/ecmlpkdd/.

Ferreira, C. A., Gama, J., & Costa, V. S. (2011, October). Constrained sequential pattern knowledge in multi-relational learning. In *Portuguese Conference on Artificial Intelligence (pp. 282-296). Springer,* Berlin, Heidelberg.

Gadekallu, T. R., Khare, N., et al. (2020). Deep neural networks to predict diabetic retinopathy. *Journal of Ambient Intelligence and Humanized Computing,* 1-14.

Hashem, S., Esmat, G., et al. (2017). Comparison of machine learning approaches for prediction of advanced liver fibrosis in chronic hepatitis C patients. *IEEE/ACM transactions on computational biology and bioinformatics, 15(3)*, 861-868.

Larranaga, P., Calvo, B., et al. (2006). Machine learning in bioinformatics. *Briefings in bioinformatics, 7*(1), 86-112.

Li, F., Gao, X., & Wan, K. (2018). Training restricted boltzmann machine using gradient fixing based algorithm. *Chinese Journal of Electronics, 27*(4), 694-703.

Liao, M., Zhao, Y. Q., et al. (2017). Automatic liver segmentation from abdominal CT volumes using graph cuts and border marching. *Computer methods and programs in biomedicine, 143, 1-12.*

Mahmoud, A. M., Alrowais, F., & Karamti, H. (2020). A hybrid deep contractive autoencoder and restricted boltzmann machine approach to differentiate representation of female brain disorder. *Procedia Computer Science, 176, 1033-1042.*

Min, S., Lee, B., & Yoon, S. (2017). Deep learning in bioinformatics. *Briefings in bioinformatics, 18*(5), 851-869.

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics, 19*(6), 1236-1246.

Movahedi, M. M., Zamani, A., et al. (2020). Automated analysis of ultrasound videos for detection of breast lesions. *Middle East Journal of Cancer, 11*(1), 80-90.

Nilashi, M., Ahmadi, H., Shahmoradi, L., Ibrahim, O., & Akbari, E. (2019). A predictive method for hepatitis disease diagnosis using ensembles of neuro-fuzzy technique. *Journal of infection and public health, 12*(1), 13-20.

Nunavath, V., Goodwin, M., Fidje, J. T., & Moe, C. E. (2018, September). Deep neural networks for prediction of exacerbations of patients with chronic obstructive pulmonary disease. In *International Conference on Engineering Applications of Neural Networks (pp. 217-228).* Springer, Cham.

Podlaha, O., Revill, P., et al. (2017). Whole-genome deep sequencing of hepatitis B virus in chronic hepatitis B patients reveals single nucleotide variants associated with baseline HBV DNA levels and HBeAg status. *Journal of Hepatology, 1*(66), S679.

Quer, J., Rodríguez-Frias, F., et al. (2017). *Deep sequencing in the management of*

hepatitis virus infections. Virus research, 239, 115-125.

Razavi, H. (2020). Global epidemiology of viral hepatitis. Gastroenterology *Clinics, 49(2), 179-189.*

Remita, M. A., Halioui, A., et al. (2017). A *machine learning approach for viral genome classification. BMC bioinformatics, 18*(1), 1-11.

Schulte, O., Bina, B., et al. (2013, April). A hierarchy of independence assumptions for multi-relational Bayes net classifiers. In *2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)* (pp. 150-159). IEEE.

Sia, F., Alfred, R., & Chin, K. O. (2013, August). Learning relational data based on multiple instances of summarized data using DARA. In *International Multi-Conference on Artificial Intelligence Technology (pp. 293-301).* Springer, Berlin, Heidelberg.

Sun, T., Zhou, B., Lai, L., & Pei, J. (2017). Sequence-based prediction of protein protein interaction using a deep-learning algorithm. *BMC bioinformatics, 18*(1), 1-8.

Taherkhani, A., Cosma, G., & McGinnity, T. M. (2018). Deep-FS*: A feature selection algorithm for Deep Boltzmann Machines. Neurocomputing, 322,* 22-37.

Upadhya, V., & Sastry, P. S. (2019). An *overview of restricted Boltzmann* machines. *Journal of the Indian Institute of Science, 99*(2), 225-236.

Varsamou, M., & Antonakopoulos, T. (2019, September). Classification using Discriminative Restricted Boltzmann Machines on Spark. In *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM) (pp. 1-6).* IEEE.

Wang, C., Tan, X. P., Tor, S. B., & Lim, C. S. (2020). Machine learning in additive manufacturing: State-of-the-art and perspectives. *Additive Manufacturing, 36,* 101538.

Wani, M. A., Bhat, F. A., Afzal, S., & Khan, A. I. (2020). Advances in deep learning. Springer.

World Health Organization. (2017). Regional action plan for the implementation of the global health sector strategy on viral hepatitis 2017–2021 (No. WHO-EM/STD/188/E). World Health Organization. *Regional Office for the Eastern Mediterranean.*

Xiao, Y., Wu, J., Lin, Z., & Zhao, X. (2018). A deep learning-based multi-model ensemble method for cancer prediction. *Computer methods and programs in biomedicine, 153,* 1-9.

Xiao, Y., Xing, C., Zhang, T., & Zhao, Z. (2019). An intrusion detection model based on feature reduction and convolutional neural networks. *IEEE Access, 7,* 42210-42219.

Zhang, Y., Li, P., & Wang, X. (2019). Intrusion detection for IoT based on improved genetic algorithm and deep belief network. *IEEE Access, 7,* 31711-31722.

Submitted: 23.02.2022
Accepted: 20.05.2022