

Received August 12, 2021, accepted September 7, 2021, date of publication September 21, 2021, date of current version October 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3114306

Early Prediction of Chronic Kidney Disease Using Deep Belief Network

SHAHINDA MOHAMED MOSTAFA ELKHOLY^{ID}, AMIRA REZK^{ID},
AND AHMED ABO EL FETOH SALEH

Department of Information System, Faculty of Computers and Information, Mansoura University, Mansoura 35516, Egypt

Corresponding author: Shahinda Mohamed Mostafa Elkholy (shahinda_alkholi@mans.edu.eg)

ABSTRACT Chronic kidney disease (CKD) is still a health concern despite advances in surgical care and treatment. CKD's growth in recent years has gained much interest from researchers around the world in developing high-performance methods for diagnosis, treatment and preventive therapy. Improved performance can be accomplished by learning the features that are in the concern of the problem. In addition to the clinical examination, analysis of the medical data for the patients can help the health care partners to predict the disease in early stage. Although there are many tries to build intelligent systems to predict the CKD by analysis the health data, the performance of these systems still need enhancement. This Paper proposes an intelligent classification and prediction model. It uses modified Deep Belief Network (DBN) as classification algorithm to predict the kidney related diseases and the Softmax as activation function and the Categorical Cross-entropy as a loss function. The evaluation of the proposed model shows that the model can predict the CKD with accuracy 98.5% and sensitivity 87.5 % comparing with existing models. Result analysis proves that using advanced deep learning techniques is beneficial for clinical decision making and can aid in early prediction of CKD and its related phases that reduces the progression of the kidney damage.

INDEX TERMS Categorical cross-entropy, chronic kidney disease, deep belief network, deep learning, softmax.

I. INTRODUCTION

Kidney disease has become a typical disease with severe problems [1]. Kidney disease is described as a heterogeneous cluster of disorders affecting the structure and performance of the urinary organ. It is widely known that even delicate anomalies in measures of the composition and quality of urinary organs are still associated with increased risk of complications in alternate organ systems [2].

There are four stages of kidney diseases:

1. NO Kidney Disease (NKD)
2. Chronic Kidney Disease (CKD)
3. Acute Kidney Injury (AKI)
4. End Stage Kidney Disease (ESKD)

Around 10% of the inhabitance in the world suffers from chronic kidney ailment (CKD), leading to millions of deaths every year [3].

The main factors specified as a reason for the disease are [4]:

The associate editor coordinating the review of this manuscript and approving it for publication was Juntao Fei^{ID}.

- Obesity and hypertension: Many medical conditions can cause KD.
- Family history: If anyone in your family has kidney disease, dialysis, or kidney transplantation, you may be more likely to develop kidney disease than someone without this family history.
- Medicines: Some medicines can cause or exacerbate kidney disease, such as over-the-counter pain medicines.
- Age and race: older people and certain racial groups may have a higher chance of developing renal disease.

The diagnosis of kidney disease in early stage saves the patient from serious complications. To predict the kidney diseases, the factors that cause it must be studied carefully.

All these factors can be translated into data to predict kidney disease and suggest a medical protocol to improve the patient health state. Medical information includes continuity, multi-attribution, incompleteness, and time-related characteristics. The problem of using large volumes of data efficiently is becoming a major issue for the healthcare industry [5].

Nowadays, in health care domain, data mining plays an indispensable role in disclosing anonymous and valuable knowledge in health data. Health care partners can ameliorate

quality of service by identifying latent, potentially valuable patterns that medical diagnosis demands by applying data mining techniques to health care domain [6].

Classification is one of healthcare organization's most utilized forms for data mining. The classification method determines the target category for each set of data. Classification is a valuable data mining process that suggests classification system (called classifier) to new cases.

Recently, in the domain of kidney disease, Deep Learning (DL) methods are used to automate the extraction and interpretation of the functions, hence models based on these techniques achieve high performance [7].

Deep learning is a machine learning focuses on learning to describe and construct multiple levels. Deep Neural Network (DNN) is an input-output neural network with multiple layers of nodes. It Identifies and processes series of layers between input and output in several stages. Deep learning requires computational models consisting of multiple layers of storage to acquire data structures of different abstraction scales. Deep learning approaches aim to learn features in hierarchical structures where features at higher hierarchy levels are created by lower-level features composition [8].

This paper introduces recent literature review in section 2, section 3 discusses the base methods and materials. The experiment design, setup, metrics and results are presented in section 4 and finally the conclusion.

II. LITERATURE REVIEW

Many researchers interested in prediction CKD. They used different classification algorithm to get an efficient and accurate prediction system.

Dutta and Bandyopadhyay [9] proposed a Neural Network with 10 folds cross validation methodology to classify the CKD patients comparing that technique with Decision tree, Support Vector Machine, Gradient Boost classifier and K Nearest Neighbours. Experimental results based on accuracy, f1 score, kappa score and MSE showed that neural network with 10 folds cross validation model performed the highest accuracy.

Zhang *et al.* [10] proposed a new approach for the transfer learning based using DenseNet-201, DenseNet-121, Densenet-169 and neural networks, and compared between them for multiple sclerosis (MS) classification. Histogram stretching (HS) was utilized for preprocessing all used images and also assigned different composite learning factor (CLF) to the three layers used. The experimental results showed that the proposed model of DenseNet-201-D is the best one comparing with state-of-the-art approaches.

Anupama *et al.* [11] developed an intracerebral hemorrhage (ICH) diagnosing model based on GrabCut segmentation technique with using of synergic deep learning (SDL) for the feature extraction, also that model used Gabor filtering for removing the noise to raise the quality of the image.

Gupta *et al.* [12] built a deep learning classification framework using the stacked autoencoder model to extract the most effective and useful features of the chronic kidney disease

dataset and the softmax classifier for prediction, so it provided high accuracy.

Rodrigues *et al.* [13] established a hybrid model of Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN), to provide high defined captions and visual descriptions for video images. They used the CNN to extract the features of the video images then, these features were utilized by Long Short-term memory (LSTM) to generate sentences with a meaningful sequence and were trained with SoftMax function.

Sharma *et al.* [14] developed an expert system to help the skin specialists using Residual Neural Networks (ResNet) with 50-layers for diagnosing and identifying various skin diseases.

Wang *et al.* [15] proposed an AlexNet model for alcoholism identification that radiologists can use in patient diagnosis. The AlexNet was used as the main transfer learning model and applied five techniques of data augmentation (DA): random translation, gamma correction, noise injection, scaling and image rotation that improve the performance of classification.

Zhao *et al.* [16] established and validated the eGFR prediction model using information derived from the regional health network. The model was developed using regression from Random Forest and tested using statistics on goodness of fit and measures of discrimination. The eGFRs estimates were used to identify patients with elevated macro-averaged and micro-averaged parameters into CKD levels.

Abdelaziz *et al.* [17] proposed a CKD-based Cloud-IOT hybrid intelligent model utilizing two intelligent methods, Linear Regression (LR) and Neural Network (NN). LR is used to identify critical factors affecting CKD while NN is used for prediction. Furthermore, as an instance of a cloud technology environment, a hybrid cognitive framework was implemented on windows azure to forecast CKD and help clinicians in smart cities.

Chaitanya and Rajesh *et al.* [18] attempted to combine various algorithms such as (ANN+GSA), (ANN+GA) and the K-nearest neighbor. Comparing the performance of these algorithms based on accuracy, sensitivity, and specificity showed that, GA-ANN algorithms have the better classification effectiveness.

Besra and Majhi [19] started with preprocessing, then grouped instances into CKD or NOTCKD, and the percentage of GFR stage calculated. The methodology for classification started with the introduction of different classifiers such as Naive-Bayes, SOM, IB1, VFI, Multi-classifier, and Random Forest. Eventually, the rate and phases of kidney activity were assessed using the GFR test method.

Tekale *et al.* [20] built two different ML models using Decision Tree and SVM to predict CKD. The comparison between them, based on variety of evaluation measures, showed that the SVM performed better with accuracy of 96.75%.

Gharibdousti *et al.* [21] used Decision Tree, Linear Regressing, Super Vector Machine, Naive Bayesian and

Neural Network as classification strategies. They investigated the similarity of the characteristics to achieve a matrix of correlation.

Boukenze *et al.* [22] Predicted kidney disease using multiple learning algorithms like SVM, MLP, DT (C4.5), Bayesian Network and KNN. They compared these algorithms and defined the most efficient algorithm based on multiple criteria; results showed that MLP and C4.5 have the best rates of accuracy.

Jojoa *et al.* [23] combined the elements used for (CKD) recognition. Two major factors, including prognosis, were modern ways of research in systematic understanding of the conditions, and allowed to develop a better and more accurate treatment plan by providing quantifiable prognostic risk factors.

Arora and Sharma [24] used three classification algorithms to detect CKD at the earliest possible stage, Naïve Bayes, J48 and SOM. Experiments with these algorithms on these datasets were implemented using WEKA, a tool of data mining. Ultimately, J48 was reported as the best algorithm for reliable and early diagnosis of chronic kidney disease. It achieved the highest accuracy among the three early-stage algorithms.

Başar *et al.* [25] proposed a chronic kidney disease diagnosis using learning algorithm of Adaboost ensemble. In this diagnostic approach, decision tree-based classifiers were used. Parameters such as: mean absolute error (MAE), kappa, root mean squared error (RMSE) and area under curve (AUC) were used to determine the category efficiency. The result showed that adaboost ensemble training algorithm delivered better output in classification than human classification.

Eroğlu and Palabaş [26] introduced six different basic classifiers, namely: K Nearest Neighbor (KNN), Naive Bayes, Support Vector Machines (SVM), Random Trees, J48, Decision Tables, and three different ensemble algorithms, namely: bagging, Ada Boost, random subspace. The result showed—that J48-based algorithm with random subspace and bagging ensemble algorithms, and random tree-based algorithm with bagging ensemble algorithm had achieved 100% classification efficiency.

Kunwar *et al.* [27] predicted chronic kidney disease (CKD) by techniques means of classification such as the Artificial Neural Network (ANN) and Naive Bayes. Rapid miner software was used for experimental purposes and the results showed that Naive Bayes was producing more accurate results than the Artificial Neural Network as it taken little time.

Ani *et al.* [28] established and created a system for promoting medical decision-making using Classification methods such as probability-based Naive Bayes, Neural Network-based Back Propagation (BPN), LDA classification, node-based decision tree, lazy learner K Nearest Neighbor (KNN) and algorithms for random subspace classification were studied. The accuracy of these algorithms were 78%, 81.5%, 90%, 93% 76% and 94% respectively on the UCI dataset.

Although the previous work did great efforts, the performance of their prediction systems still need enhancement to get high accurate results. So, this paper will propose a model to predict CKD with highest performance based on deep learning.

III. PROPOSED MODEL

The methodology of our research consists of sequence steps including data collection, pre-processing, deep learning-based feature engineering and classification followed by evaluation. These computational steps are graphically presented in Figure 1.

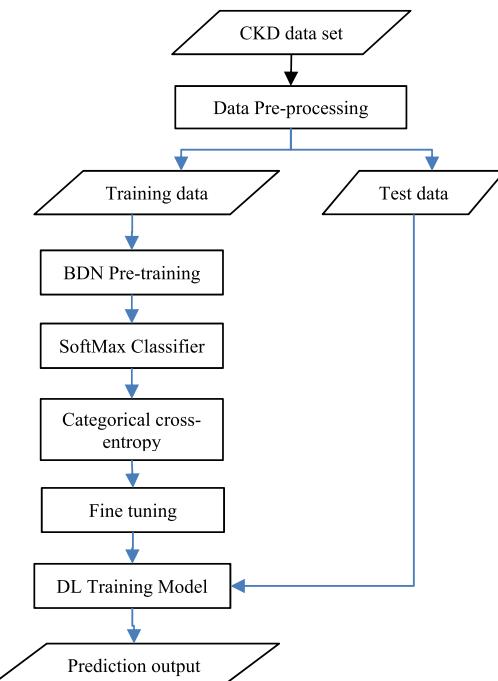


FIGURE 1. Workflow diagram for the proposed deep belief network.

A. CKD DATA SET

The proposed model uses UCI Dataset [29]. It consists of 25 attributes, 14 nominal, 11 numeric and 1 class attribute which represent the data of 400 patients out of which 250 states are CKD and 150 are not. Table 1 shows the description of the used dataset.

B. DATASET PRE-PROCESSING

Classifying data with missing values is a challenge.

The used dataset has missing values, which reduce the efficiency, so it must be removed before analyzing data.

The missing values can be determined in two points of view, cases (records) or attributes.

In cases(record) point of view, the missing values degree may be simple, medium, or complex. It is simple degree if the case (record) has a missing value in one attribute at most. It is medium if the case (record) has missing values in 2% to 50% of the total number of attributes. While It is complex if the case (record) has missing values in at least 50% up to 80% of attributes.

TABLE 1. Dataset description of CKD.

ID	Attribute Symbols	Attribute Description	Attribute Type
1	Age	Numerical	Years
2	Blood pressure	Numerical	Mm/Hg
3	Specific gravity	Nominal	0.005,1.010,1.015,1.020,1.025
4	Albumin	Nominal	0.1,2,3,4,5
5	Sugar	Nominal	0.1,2,3,4,5
6	Red blood cells	Nominal	Normal, abnormal
7	Pus cells	Nominal	Normal, abnormal
8	Pus cell clumps	Nominal	Present, not present
9	Bacteria	Nominal	Present, not present
10	Blood glucose	Random	Mgs/dl
11	Blood urea	Numerical	Mgs/dl
12	Serum creatinine	Numerical	Mgs/dl
13	Sodium	Numerical	mEq/L
14	Potassium	Numerical	mEq/L
15	Hemoglobin	Numerical	Gms
16	Packed cell volume	Numerical	Gms
17	White blood cell count	Numerical	Cells/cum
18	Red blood cell count	Numerical	Millions/cm
19	Hypertension	Nominal	Yes, no
20	Diabetes mellitus	Nominal	Yes, no
21	Coronary artery disease	Nominal	Yes, no
22	Appetite	Nominal	Good, poor
23	Pedal edema	Nominal	Yes, no
24	Anemia	Nominal	Yes, no
25	Class	Nominal	CKD, NOCKD

In the attributes point of view, the missing values may be univariate, monotone, or arbitrary. It is univariate if all missing values in one attribute (feature). It is monotone if the missing values are in at least three attributes. While it is arbitrary if the missing values are in random attributes [30].

Imputation is seen as a way to avoid pitfalls of cases with missing values.

The reason behind missing value imputation is that dataset is not decreased, because the data imputation retains the entire sample size. There are many methods of imputations with different features to impute the missing data.

The mode imputation is used for nominal attributes which takes the most frequent value per column and for the numerical attributes the median imputation is used in which using the mid-range value of the attribute. Some of missing values are shown in Figure 2.

Figure 2 shows that the data set contains complex and arbitrary missing values which will be overcome in the pre-processing stage.

C. DEEP LEARNING TRAINING

We performed our experiments on DBN based architectures to train and evaluate the dataset. DBN is based essentially on Restricted Boltzmann Machine (RBM) [31].

1) RESTRICTED BOLTZMAN MACHINE

RBM is a two-layer bipartisan graphic design with a collection of visible unit v , a set of invisible units h , and symmetrical relations defined by a weight matrix W between these two layers. The joint probabilities of the RBM of the hidden units

Index	age	bp	sg	al	su	rbc	pc	pt	ba	bgr	bri	na	smo	
0	48	98	1.45	1	0	7	normal	notpresent	121	56	12.9	7	7	
1	3	98	1.42	4	0	7	normal	notpresent	121	58	4.6	7	7	
2	82	98	1.41	2	0	7	normal	notpresent	123	53	11.6	7	7	
3	82	98	1.405	2	0	7	normal	notpresent	127	56	3.8	113	7	
4	91	98	1.41	2	0	7	normal	notpresent	106	28	3.4	7	7	
5	88	98	1.415	3	0	7	normal	notpresent	74	25	5.1	142	7	
6	68	79	1.41	0	0	7	normal	notpresent	100	54	24	184	7	
7	52	100	1.415	3	0	7	abnormal	present	138	68	1.9	7	7	
8	53	98	1.42	2	0	7	abnormal	present	78	187	7.2	114	7	
9	58	98	1.41	2	4	7	abnormal	present	498	55	4	7	7	
10	63	70	1.41	3	0	7	abnormal	present	368	68	2.7	131	7	
11	68	70	1.415	3	1	7	normal	present	288	72	2.1	138	7	
12	68	70	1.41	7	7	7	?	notpresent	98	86	4.6	135	7	
13	68	98	1.405	3	2	7	normal	abnormal	present	157	98	4.3	135	7
14	68	98	1.405	3	0	7	?	?	?	162	9.6	141	7	7
15	87	79	1.415	2	0	7	normal	notpresent	98	46	2.7	128	7	
16	47	98	1.415	7	7	7	?	?	?	124	87	5.2	139	7
17	68	100	1.425	0	3	7	normal	notpresent	263	27	1.3	135	7	
18	62	98	1.415	1	0	7	abnormal	present	198	31	5.6	7	7	
19	81	98	1.415	2	0	7	abnormal	notpresent	149	54	3.9	135	7	
20	68	98	1.41	7	7	7	?	?	?	168	76	4.5	7	7
21	48	98	1.425	4	0	7	normal	abnormal	notpresent	95	163	7.7	136	7
22	42	100	1.415	4	0	7	normal	abnormal	present	1	58	1.4	129	7
23	61	68	1.425	0	7	?	?	?	?	180	75	1.9	141	7
24	69	98	1.415	0	0	7	normal	notpresent	156	45	2.4	148	7	
25	69	70	1.41	3	4	7	normal	notpresent	264	87	2.7	138	7	

FIGURE 2. Data with missing values.

h and visible units v is described by:

$$p(v, h; W, a, b) = \frac{1}{Z} e^{E(v, h; W, a, b)} \quad (1)$$

where E is the energy function, Z is the partition function, while W , a , and b are model parameters. The partition function Z is defined by:

$$Z = \sum_{v', h'} e^{E(v', h'; W, a, b)} \quad (2)$$

The energy function E is:

$$E(v, h; W, a, b) = -h^T W v - b^T h - c^T v \quad (3)$$

where T represents the transpose operator.

The RBM is pre-trained by the Contrastive Divergence (CD) training algorithm using training data.

The method of CD preparation is carried out with the stochastic steepest ascending [32], as seen in Figure 3. Starting with Training Data $v0$, Equation (4) assumes the hidden state $h0$ then retrieves training data with invisible state of $h0$ by (5). Reconstruction $v1$ assume to the secret state $h1$.

The hidden state's conditional probabilities presented the visible state v are represented by:

$$P(h = 1/v) = F(W^T v + b) \quad (4)$$

The visible state is given by binary random state. The conditional probabilities of the visible state which provided the hidden state are described in similar ways.

$$P(v = 1/h) = F(Wh + c) \quad (5)$$

The output Y of a layer's nodes is defined by:

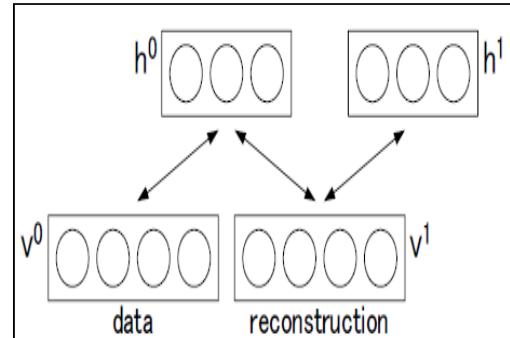
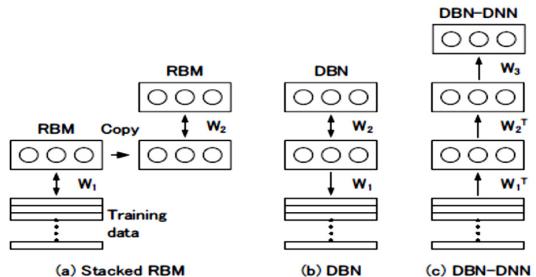
$$Y = F(W_i^T EP(x_i) [x_i] + b_i) \quad (6)$$

where x is a random variable, E is the expectation of this random variable; P shows the Probabilities of this element and F represent the used activation function [31].

The algorithm of CD for RBM is shown in algorithm 1 [33].

Algorithm 1 k-Step Contrastive Divergence for RBMs

Input: RBM ($V_1, \dots, V_m, H_1, \dots, H_n$), training period T , learning rate ϱ
Output: The RBM weight matrix ω , gradient approximation ω_{ij} , a_i and b_j for $i = 1, \dots, m, j = 1, \dots, n$
 Initialize ω with random values distributed uniformly in $[0, 1]$, $\omega_{ij} = a_i = b_j = 0$ for $i = 1, \dots, m, j = 1, \dots, n$
 For $\forall t = 1: T$ do,
 Sample $v_i^{(t)}$
 $P(v_i|h^{(t)})$ when $i = 1, \dots, m$
 Sample $h_j^{(t+1)}$
 $P(h_j|v^{(t)})$ when $j = 1, \dots, n$
 For $i = 1, \dots, m, j = 1, \dots, n$ do
 $\Delta\omega_{ij} = \Delta\omega_{ij} + \varrho \times [P(H_j = 1|v^{(0)}) \times v_i^{(0)} - P(H_j = 1|v^{(T)}) \times v_i^{(T)}]$
 $\Delta a_i = \Delta a_i + \varrho \times (v_i^{(0)} - v_i^{(T)})$
 $\Delta b_j = \Delta b_j + \varrho \times [P(H_j = 1|v^{(0)}) - P(H_j = 1|v^{(T)})]$
 $\omega = \omega + \varrho \times \Delta\omega_{ij}$
 End For
 End For

**FIGURE 3.** Contrastive divergence algorithm for RBM.**FIGURE 4.** Model architecture of deep belief network classifier.**2) DEEP BELIEF NETWORKS (DBN)**

Deep Belief Networks are interactive systems that built on stacking RBM that trained with CD. The algorithm that determines the optimum locale for each layer and the next stacked RBM layer takes those optimally trained values and searches for the optimum locale again that is the cause of the greedy algorithm for learning works of DBN training layer by layer as shown in Algorithm 2 [26].

Algorithm 2 Greedy Layer-by-Layer Algorithm for DBN

Input: Input visible vector v_{in} , training period T , learning rate ϱ , number of layers J .
Output: Weight matrix ω_i of layer i , $i = 1, 2, \dots, J-2$. Initialize ω with random values from 0 to 1, $h^0 = v_{in}$; where h^0 denotes the value of units in the input layer.
 h_i represents the units' value of the i th layer. $layer = 1$;
 for $\forall t = 1: T$ do,
 for layer $< L$, do gibbs sampling h^{layer} using $P(h^{layer}|h^{layer-1})$
 Computing the CD in Algorithm 1, $\omega(layer)$ is achieved using
 $\omega(t+1) = \omega(t) + \varrho \times \Delta\omega$
 End for
 End for

DBN learn to derive the training data from a deep hierarchical representation. It is a generational graphic model with multiple layers of deep-seated variables, connecting layers but not units within each layer. DBN is an unsupervised learning method, as opposed to neural networks for vision and back propagation [32]. Most current deep models start with random parameters and then learn gradually to approximate the local optimal ones. Unfortunately, weak initial parameters will contribute to a poor local optimum, which can have a significant impact on the subsequent learning method. A DBN's training process can solve this problem in the pre-training phase in which an unsupervised learning-based training is performed for setting the initial values of the weight of the network by training not randomly [33].

A simple network of beliefs [32] consists of layers of stochastic binary units with weighted relations as shown in Figure 4.

3) PROPOSED CKDPM

The proposed chronic kidney disease prediction model has deep network that is consisted of 6 layers. The first layer is for the input and this layer is constructed by 24 nodes while the second layer takes its input from the first layer; second layer is constructed with 13 nodes. Third layer consists of 28 nodes and its inputs come from the second layer. 8-nodes are used with forth layer and these nodes take the inputs from the third layer. Fifth layer consist of 4 nodes take its inputs from the fourth layer. Finally, the output layer has two nodes as it works for binary classification. Figure 5 shows the structure of the proposed deep neural network.

The model passes through three stages:

The first stage is the pre-training phase in which the input layer receives the data and with the second layer forms a typical RBM using training for reaching the local optimal parameters. Then this output is taken as the third layer's input and the third layer with the second layer form a new RBM and the parameters are trained with sing unsupervised learning method continually, RBMs appear to be greedily piled layer by layer to reach the local optimal values and the approximation of the weighting factor is done in one step, resulting in a significant reduction in training time and the problem of underfitting, which is common in other networks, can also be solved during the pre-training phase.

The second stage is the fine-tuning in which using the Back Propagation algorithm to fine-tune the whole network parameters to reach the global optimal using the up-down

algorithm, which is a contrastive variant of the wake-sleep algorithm that is used to train the DBNs with labeled data.

In the learning process of recognizing weights, a collection of labels is applied to the top layer in order to determine the network's category boundaries. The up-down algorithm, unlike the current wake-sleep algorithm, does not suffer from mode-averaging problems, which can result in low recognition weights.

The third stage is the classification (test) in that the trained model with optimal weights with biases is obtained, so data can be classified into the proper class.

Here, in our model, we use the modified DBN as in [32], which apply the DBN that built on RBM as discussed previously but it takes the expectation of the output after the activation function as in (7). Taking the expectation as late as possible is generally good, so the output Y of a layer's nodes is defined by:

$$Y = EP(x_i) [F(W_i^T [x_i] + b_i)] \quad (7)$$

The difference between (6) and (7) is where the expectation is taken.

As mentioned above, F is the activation function used.

Here we use the SoftMax activation function as denoted at (8) it estimates the exponential of the given input x_i and the summation of the exponential of all input values and the fraction of these results in the SoftMax function output [34].

The output of the SoftMax function is the probabilities of each class and the class of the highest probability is the output class.

$$\hat{y} = F(x_i) = \frac{\text{Exp}(x_i)}{\sum_{j=0}^k \text{Exp}(x_j)} \quad \text{where } i = 0, 1, \dots, k \quad (8)$$

Finally, in our model, the Categorical Cross-entropy is a loss function that used for single label identification. It calculates the square difference between the truth output and the predicted one as in (9).

$$L(y, \hat{y}) = \sum_{i=0}^m \sum_{j=0}^n (y_{ij} * \log(\hat{y}_{ij})) \quad (9)$$

where \hat{y} is the predicted output value. After applying the activation function in the output layer, the Categorical cross-entropy will compare the predictions with the true values [35].

In the fine-tuning the updated weights are propagated to the hidden and visible unites by the back propagation algorithm. For the final layer updates will be as follows:

$$\frac{\partial L}{\partial W_{ij}} = \delta_j \frac{\partial \mu_j}{\partial W_{ij}} + T_j \frac{\partial \rho_j^2}{\partial W_{ij}} \quad (10)$$

$$\delta_j = \frac{\partial L}{\partial \mu_j} = \frac{\partial L}{\partial o_j} \frac{\partial o_j}{\partial n_j} \frac{\partial n_j}{\partial \mu_j} \quad (11)$$

$$T_j = \frac{\partial L}{\partial \rho_j^2} = \frac{\partial L}{\partial o_j} \frac{\partial o_j}{\partial n_j} \frac{\partial n_j}{\partial \rho_j^2} \quad (12)$$

where L is the loss function, o_j is the j-th node output, n_j is the activation function's input of j -th node, i and j identify the input and the output nodes.

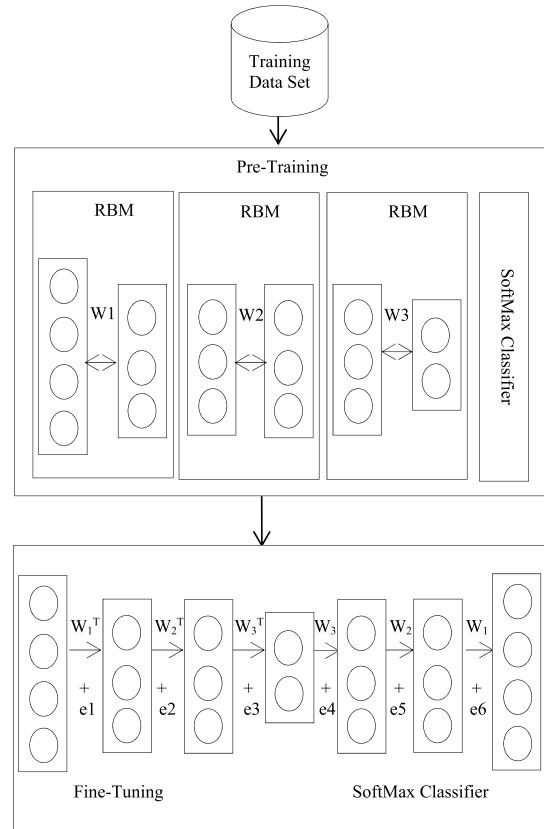


FIGURE 5. The proposed deep model. The model has 3 stages pre-training, the classification, and the fine-tuning.

For the hidden layer, the updates will be:

$$\frac{\partial L}{\partial W_{ij}} = \delta_j \frac{\partial \mu_j}{\partial W_{ij}} + T_j \frac{\partial \rho_j^2}{\partial W_{ij}} \quad (13)$$

$$\delta_j = \alpha_j \frac{\partial n_j}{\partial \mu_j}, \quad (14)$$

$$T_j = \alpha_j \frac{\partial n_j}{\partial \rho_j^2} \quad (15)$$

$$\alpha_j = [\sum_k \{\delta_k \frac{\partial \mu_k}{\partial o_j} + T_k \frac{\partial \rho_k^2}{\partial o_j}\}] \frac{\partial o_j}{\partial n_j} \quad (16)$$

where δ_k and T_k are propagated updates from the previous layer, and k identify the previous layer's node [32].

4) MATERIALS (SOFTWARE AND HARDWARE) USED

IBM core i7 processor computer with 16 GB DDRAM and NVIDIA GeForce MX150 graphics card is used to conduct tests and evaluations. The coded program is implemented using MATLAB 2018 x64-bit.

5) MODEL EVALUATION

The dataset is randomly divided in two parts; the first part contains 70% of overall collection of data to train the model. The second part is used for testing and it contains the reminder of the dataset (30%). Six performance measures are used to evaluate and validate the proposed model. These measures are Accuracy, Precision, Recall, F-Measure [36],

TABLE 2. Proposed model performance.

Measures	Training	Testing
Accuracy	98.52	98.51
Precision	89.95	86.82
Recall	87.70	87.50
F-measure	87.70	87.50
RMSE	0.484552	0.482874
Mean Absolute Error	0.376666	0.370000

TABLE 3. Accuracy of the proposed model comparing with other models.

	Used Techniques	Accuracy	F1 Measure
Milandeep Arora and Er. Ajay Sharma 2016 [24]	Naïve Bayes	95.556%	0.94
B. Boukenze et al. 2017 [22]	J48	97.86%	0.97
	SMO	97.75%	0.96
S. M. K. Chaitanya et al. 2019 [18]	SVM	60.25	0.97
	MLP	62.25	0.99
	DT (C4.5)	63%	0.98
	Bayesian Network	57.5%	0.93
	KNN	58.25 %	0.94
Siddheshwar Tekale et al. 2018 [20]	Decision Tree	91.75 %	0.89
	SVM	96.75 %	0.95
Shawni Dutta,SamirKumar 2020 [9]	(ANN+GSA)	78%	0.96
	(ANN+GA)	56%	0.93
	KNN	67%	0.975
Proposed Model	Deep Believe Network (DBN)	98.52 %	0.87

Root Mean Square Error (RMSE) [37] and Mean Absolute Error (MAE) [38]. Next equations show how these measures computed:

$$\text{Accuracy} = \frac{t_p + t_n}{t_p + f_p + f_n + t_n} \quad (17)$$

$$\text{Precision} = \frac{t_p}{t_p + f_p} \quad (18)$$

$$\text{Recall} = \frac{t_p}{t_p + f_n} \quad (19)$$

$$F - \text{measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (20)$$

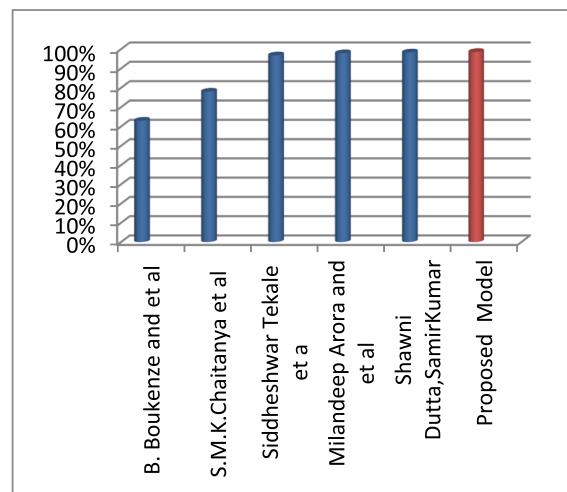
$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_t - y_t)^2}{n}} \quad (21)$$

$$\text{MAE} = \frac{|\hat{y}_t - y_t|}{n} \quad (22)$$

where t_p, f_p, f_n , and t_n refer to true positive, false positive, false negative, and true negative respectively [35].

IV. RESULTS AND DISCUSSIONS

Table 2 shows the obtained results for the proposed model during training and testing. Figure 6 and Table 3 show

**FIGURE 6.** Accuracy of the proposed model comparing with other models using the same data set.

comparison of our model with others which used the same dataset but using different algorithms based on various evaluation measures. The dataset is balanced so the accuracy is perfect and is the most important measure on which all the classifiers depend on, that shows the correctly classified instances percentage. The proposed model shows high accuracy because of the advantages of the DBN that per-training the model in which the values of the local optimal and weighting factors are done in one step. It overcomes the underfitting problem and fine-tunes the whole deep model in one iteration, So the training time is reduced.

Finally, the results of comparison revealed that the proposed model is effective in classification of CKD dataset.

V. CONCLUSION

This research presents an intelligent model of deep neural network for prediction and classification of chronic kidney disease dataset using a deep belief network. The model is built by using DBN with SoftMax classifier and the Categorical Cross-entropy as a loss function. We use dataset from UCI's machine learning database and perform preprocessing to handle the missing values.

The efficiency of the proposed model is evaluated, and also made a comparison with the existing models. The proposed model achieves better performance, comparing with the existing models, with accuracy 98.52%. So, the proposed model presents proper predictor and classifier for CKD.

REFERENCES

- [1] [Online]. Available: <https://www.kidney.org/sites/default/files/11-10-0101>
- [2] [Online]. Available: [https://www.ajkd.org/article/S0272-6386\(13\)00590-8/fulltext](https://www.ajkd.org/article/S0272-6386(13)00590-8/fulltext)
- [3] [Online]. Available: <https://www.asn-online.org/kidneydisease/info/PEM-text.pdf>
- [4] [Online]. Available: <https://www.researchgate.net/publication/329521615>
- [5] H. C. Koh and G. Tan, "Data mining applications in healthcare," *J. Healthcare Inf. Manag.*, vol. 19, no. 2, pp. 64–72, 2005.
- [6] M. Bedford, C. Farmer, A. Levin, T. Ali, and P. Stevens, "Acute kidney injury and CKD: Chicken or egg?" *Amer. J. Kidney Diseases*, vol. 59, no. 4, pp. 485–491, Apr. 2012.

- [7] H. Kriplani, B. Patel, and S. Roy, "Prediction of chronic kidney diseases using deep artificial neural network technique," in *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*, 2019, pp. 18–179, doi: [10.1007/978-3-030-04061-1_18](https://doi.org/10.1007/978-3-030-04061-1_18).
- [8] T. Karthikeyan and V. A. Kanimozh, "Deep learning approach for prediction of heart disease using data mining classification algorithm deep belief network," *Int. J. Adv. Res. Sci. Eng. Technol.*, vol. 4, no. 1, pp. 1–8, 2017.
- [9] S. Dutta and S. K. Bandyopadhyay, "Chronic kidney disease prediction using neural approach," Tech. Rep., Jun. 2020, doi: [10.1101/2020.06.28.20142034](https://doi.org/10.1101/2020.06.28.20142034).
- [10] [Online]. Available: <https://www.researchgate.net/publication/342356888>
- [11] C. S. S. Anupama, M. Sivaram, E. L. Lydia, D. Gupta, and K. Shankar, "Synergic deep learning model-based automated detection and classification of brain intracranial hemorrhage images in wearable networks," *Pers. Ubiquitous Comput.*, Nov. 2020, doi: [10.1007/s00779-020-01492-2](https://doi.org/10.1007/s00779-020-01492-2).
- [12] A. Khamparia, G. Saini, B. Pandey, S. Tiwari, D. Gupta, and A. Khanna, "KDSAE: Chronic kidney disease classification with multi-media data learning using deep stacked autoencoder network," *Multimedia Tools Appl.*, vol. 79, nos. 47–48, pp. 35425–35440, Dec. 2020, doi: [10.1007/s11042-019-07839-z](https://doi.org/10.1007/s11042-019-07839-z).
- [13] A. Khamparia, B. Pandey, S. Tiwari, D. Gupta, A. Khanna, and J. J. P. C. Rodrigues, "An integrated hybrid CNN-RNN model for visual description and generation of captions," *Circuits Syst. Signal Process.*, vol. 39, pp. 776–788, Nov. 2020, doi: [10.1007/s00034-019-01306-8](https://doi.org/10.1007/s00034-019-01306-8).
- [14] M. Sharma, B. Jain, C. Kargetti, V. Gupta, and D. Gupta, "Detection and diagnosis of skin diseases using residual neural networks (RESNET)," *Int. J. Image Graph.*, Dec. 2020, Art. no. 2140002, doi: [10.1142/S0219467821400027](https://doi.org/10.1142/S0219467821400027).
- [15] S.-H. Wang, S. Xie, X. Chen, D. S. Guttley, C. Tang, J. Sun, and Y.-D. Zhang, "Alcoholism identification based on an AlexNet transfer learning model," *Frontiers Psychiatry*, vol. 10, Apr. 2019.
- [16] J. Zhao, S. Gu, and A. McDermaid, "Predicting outcomes of chronic kidney disease from EMR data based on random forest regression," *Math. Biosci.*, vol. 310, pp. 24–30, Apr. 2019.
- [17] A. Abdelaziz, A. S. Salama, A. M. Riad, and A. N. Mahmoud, "A machine learning model for predicting of chronic kidney disease based Internet of Things and cloud computing in smart cities," in *Security in Smart Cities: Models, Applications, and Challenges* (Lecture Notes in Intelligent Transportation and Infrastructure), 2019, pp. 93–114, doi: [10.1007/978-3-030-01560-2_5](https://doi.org/10.1007/978-3-030-01560-2_5).
- [18] S. M. K. Chaitanya and P. R. Kumar, "Detection of chronic kidney disease by using artificial neural networks and gravitational search algorithm," in *Innovations in Electronics and Communication Engineering* (Lecture Notes in Networks and Systems), vol. 33, 2019.
- [19] Ankit, B. Besra, and B. Majhi, "An analysis on chronic kidney disease prediction system: Cleaning, preprocessing, and effective classification of data," in *Recent Findings in Intelligent Computing Techniques (Advances in Intelligent Systems and Computing)*, vol. 707, 2019.
- [20] S. Tekale, P. Shingavi, and S. Wandhekar, "Prediction of chronic kidney disease using machine learning algorithm," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 7, no. 10, pp. 92–96, Oct. 2018.
- [21] M. S. Gharibdousti, K. Azimi, S. Hathikal, and D. H. Won, "Prediction of chronic kidney disease using data mining techniques," in *Proc. Ind. Syst. Eng. Conf.*, 2018, pp. 2135–2140.
- [22] B. Boukenze, A. Haqiq, and H. Mousannif, "Predicting chronic kidney failure disease using data mining techniques," in *Advances in Ubiquitous Networking 2*. Singapore: Springer, 2017, pp. 701–712.
- [23] J. A. Jojoa, C. E. Rivera, E. M. Delgado, H. M. Casas, G. M. Rosas, C. Y. Rosero, and F. A. Montenegro, "New ABC chronic kidney disease classification," *Int. J. Nephrol. Kidney Failure.*, vol. 23, pp. 1–4, Sep. 2017.
- [24] M. Arora and E. A. Sharma, "Chronic kidney disease detection by analyzing medical datasets in weka," *Int. J. Comput. Appl.*, vol. 6, no. 4, pp. 1797–2250, Jul./Aug. 2016.
- [25] M. D. Başar, P. Sarı, N. Kılıç, and A. Akan, "Detection of chronic kidney disease by using AdaBoost ensemble learning approach," in *Proc. 24th Signal Process. Commun. Appl. Conf. (SIU)*, 2016, pp. 773–776.
- [26] K. Eroğlu and T. Palabaş, "The impact on the classification performance of the combined use of different classification methods and different ensemble algorithms in chronic kidney disease detection," presented at the Nat. Conf. Elect., Electron. Biomed. Eng. (ELECO), Dec. 2016.
- [27] V. Kunwar, K. Chandel, A. S. Sabitha, and A. Bansal, "Chronic kidney disease analysis using data mining classification techniques," in *Proc. 6th Int. Conf.-Cloud Syst. Big Data Eng. (Confluence)*, Jan. 2016, pp. 300–305.
- [28] R. Ani, G. Sasi, U. R. Sankar, and O. S. Deepa, "Decision support system for diagnosis and prediction of chronic renal failure using random subspace classification," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2016, pp. 1287–1292.
- [29] [Online]. Available: https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease
- [30] [Online]. Available: <https://www.researchgate.net/publication/312568863>
- [31] G. Hinton, "A practical guide to training restricted Boltzmann machines," UTML TR., Tech. Rep., 2010.
- [32] M. Tanaka and M. Okutomi, "A novel inference of a restricted Boltzmann machine," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 1526–1531.
- [33] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, Apr. 2017.
- [34] [Online]. Available: <https://www.researchgate.net/publication/328826136>
- [35] [Online]. Available: <https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/loss-functions>
- [36] H. M and S. M. N, "A review on evaluation metrics for data classification evaluations," *Int. J. Data Mining Knowl. Manage. Process*, vol. 5, no. 2, pp. 1–11, Mar. 2015, doi: [10.5121/ijdkp.2015.5201](https://doi.org/10.5121/ijdkp.2015.5201).
- [37] [Online]. Available: <https://peltarion.com/knowledge-center/documentation/evaluation-view/regression-loss-metrics/rmse-root-mean-squared-error>
- [38] [Online]. Available: <https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/loss-functions/mean-absolute-error>
- [39] A. Botchkarev, "Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology," *Interdiscipl. J. Inf., Knowl., Manage.*, vol. 14, pp. 45–79, 2019. [Online]. Available: <https://arxiv.org/abs/1809.03006>
- [40] C. Parker, "An analysis of performance measures for binary classifiers," in *Proc. IEEE 11th Int. Conf. Data Mining*, Dec. 2011, pp. 517–526.



SHAHINDA MOHAMED MOSTAFA ELKOLY was born in Egypt, in 1986. She received the B.S. and M.Sc. degrees in information systems from the Faculty of Computers and Information Systems, Mansoura University, Mansoura, Egypt, in 2007 and 2013, respectively. Her research interests include neural networks, artificial intelligence, web service composition, and evolutionary algorithms, cloud computing, big data, data analysis, and the Internet of Things.



AMIRA REZK was born in Egypt. She received the B.S., M.Sc., and Ph.D. degrees in information systems from the Faculty of Computers and Information Systems, Mansoura University, Mansoura, Egypt, in 2000, 2007, and 2012, respectively. She is currently an Associate Professor with the Department of Information Systems, Faculty of Computers and Information, Mansoura University. Her research interests include big data analytics, artificial intelligence, data mining database, and data security.



AHMED ABO EL FETOH SALEH is currently a Professor of information systems with the Faculty of Computers and Information, Mansoura University. He has published more than 40 papers in international and local conferences, journals, and proceedings in different fields of information technology. His research interests include numerical analysis operation research, computer architecture, computer networks, data mining, information theory, and discrete mathematics.