How to Cite:
Shadadi, E., & Alamer, L. (2022). A customized stacked dense network model for chronic kidney disease prediction. *International Journal of Health Sciences*, *6*(S5), 8911–8928.
https://doi.org/10.53730/ijhs.v6nS5.11024

A customized stacked dense network model for chronic kidney disease prediction

Ebtesam Shadadi
Department of Computer Science, College of computer Science and Information Technology, Jazan University, Jazan, Kingdom of Saudi Arabia
Email: ashedadi@jazanu.edu.sa

Latifah Alamer
Department of Information Technology and Security, College of computer Science and Information Technology, Jazan University, Jazan, Kingdom of Saudi Arabia
Email: aalamer@jazanu.edu.sa

***Abstract*---**Chronic kidney disease (CKD) is still a health concern, even though surgical care and treatment have improved. Recently, academics throughout the globe have been more interested in creating high-performance approaches for diagnosing, treating, and preventing kidney disease by being more knowledgeable about the aspects that the issue is concerned with, designed to provide better. Evaluation of patient records for patients may assist health care providers detects sickness earlier on. Several have tried to construct sophisticated algorithms that forecast CKD by analyzing health data, but their effectiveness needs improvement. An intelligence categorization and linear regression model are suggested in this paper. The kidney-related disorders are predicted using a customized stacked dense network model ($c - SDNM$). Compared to current models, the testing of the conceptual scheme reveals that it can predict CKD with 98.5% accuracy. Research suggests that utilizing sophisticated deep learning algorithms is advantageous for treatment decisions and may assist in the early diagnosis of CKD and its associated stages, reducing the development of kidney problems.

***Keywords*---**Chronic kidney disease, prediction, deep learning, feature analysis, classification.

1. Introduction

An increasingly common illness with serious complications is kidney disease. [1]. People say that kidney disease is a group of different illnesses that influences functional and structural level of the renal and hepatic organs. It is well recognized that even little deviations in the function and structure of the urinary organs might raise the likelihood of issues in other organ systems [2]. The four phases of renal disease are as follows: No kidney disease, first (NKD), 2. Chronic Kidney Disease (CKD), 3. Serious Kidney Damage (AKI);and 4. End-stage renal illness (ESKD) chronic kidney disease (CKD), which kills millions of people worldwide each year, affects around 10% of the world's population. The following primary causes are cited as contributing to the condition: 1) Hypertension and adiposity: KD may be brought on by a variety of illnesses, 2) Family history: Any family members have kidney problems, i.e. hemodialysis or transplanting the renal who does not have this genetic history, 3) Medications: Certain medications, including over-the-counter painkillers, may cause or worsen renal disease and 4) Age and race: Nephrology illness will be much more common in older adults and certain ethnic communities.

The patient is spared from significant consequences by promptly identifying a renal illness. The variables that cause renal illnesses must be properly examined to forecast them. These variables are converted into information to forecast renal illness and provide a treatment plan to enhance the patient's health. Healthcare data comprises features relating to continuity, multi-attribution, incompleteness, and temporality. The challenge of effectively employing massive amounts of data is increasingly important for healthcare sector [5]. In the modern applications, data mining is important for revealing anonymized and useful information inside health data. Health care partners may improve the customer experience by finding hidden, potentially beneficial connections that clinical issue necessitates using data mining methods in the healthcare sector [6]. Taxonomy is a popular method of data extraction in medical services.The categorization process decides which category should be used for each data set. Data mining methods like classification, which recommends a categorization scheme (called a classifier) for new instances, are very useful. Deep Learning (DL) approaches are now being utilized to automate the retrieval and understanding of the functions in the field of renal disease, leading to models made using these approaches achieving excellent results [7]. Deep learning is an algorithm focused on understanding to create and represent several layers. A Deep Neural Network (DNN) has numerous layers of nodes. Multiple layers between the input and output are identified and processed. Deep learning requires computer simulations with numerous backups to learn file systems with various levels of complexity. Deep learning techniques try to discover features in organizational forms, where higher-level elements are formed through synthesizing lower-level characteristics [8]. However, there are various disadvantages related to the prediction process [9] – [10]. To handle these issues, this research concentrates on modelling and efficient customized network models for CKD prediction by stacking the network. The major research contributions are:

1) An online available UCI dataset for CKD prediction is taken for prediction purposes where the univariate samples are considered for evaluation purposes;

2) The proposed customized stacked dense network model ($c-SDNM$) with polynomial functions is considered for performing the approximation process. Here, a matrix is formed, and the singular value decomposition is performed to transform the matrix to bias.
3) The stacked dense network model-based prediction is improved by provisioning the output of one layer to another. Thus, it helps to handle the data representation, and the learning model is used for propagation purposes.
4) The disease modalities' relationship is identified, and the higher-level features are measured to extract the robust outcomes.

The work is structured as follows: section 2 gives a wider analysis of various existing approaches; section 3 elaborates on the proposed $c-SDNM$ model for prediction purposes. The numerical outcomes are given in section 4 with graphical representation. The research summary is provided in section 5.

2. Related works

Numerous scholars are concerned with CKD prediction. They applied several categorization algorithms to create a consistent and valid predictive model. Gonsalves et al. [11] suggested a NN with 10 folds validation set approach to identify the CKD patients, contrasting that method with a DT, SVM, boosting classifiers, and k-NNs. Derived from empirical findings for accuracy, f1, kappa, and MSE, it was determined that neural networks with 10 folds of cross-validation functioned at the maximum accuracy. Song et al. [12] developed a novel strategy for transferring learning based on dense network and NN evaluated for multiple sclerosis (MS) categorization. Histogram stretching (HS) was employed to pre-process all images, and each layer was allocated a separate CLF. The research findings demonstrated that, when compared to cutting-edge methods, the suggested DenseNet-201-D framework is the most effective [13].

The diagnosis model is proposed based on the GrabCut segmentation approach with deep synergic learning (SDL) for feature extraction. Additionally, that model employed Gabor filter to eliminate noise and improve the image quality. Reddy et al. [14] made ofDL classification framework with stacked auto-encoder to determine the efficient benefits to its users in the chronic kidney disease dataset and the softmax multilayer perceptron to make accurate predictions. This framework was very accurate. Lee et al. [15] created a hybrid model of RNN and CNN to offer high-definition subtitles and visual characterizations for video images. The author employed CNN to identify the characteristics from the video, and the LSTM was trained using the SoftMax function using these properties to produce intelligible phrases. Quesada et al. [16] created an inference engine employing ResNet) with 50 levels to diagnose and classify skin disorders. Kim et al. [17] developed neural network architecture for identifying alcohol addiction that radiographers can use to diagnose patients. The AlexNet was utilized as the primary transfer learning method, and augmentation methods were used to enhance classification efficiency: random translations, histogram equalization, noise insertion, scaling, and picture rotations [18].

Vasiljeva et al. [19] made the eGFR prediction model and ensured it worked using information from the geographic network. The model is built with correlation and evaluated with statistics on how well the model fits and how well it separates the same thing. Patients with raised macro- and micro-averaged indicators were found using the eGFR calculations. The CKD-based Regression Analysis (LR) and Neural Networks was suggested by Huang et al. [20]. While NN is used to forecast, LR is used to pinpoint important variables impacting CKD. The cognition model was developed using Microsoft azure to anticipate CKD and support physicians in smart cities.Xu et al. [21] tried to combine various techniques. Prediction accuracy, specificity and sensitivity of GA–ANN techniques offer greater classification efficacy. Huling et al., [22] pre-processing were followed by grouping occurrences of the disease or not and calculating the GFR percentage. The emergence of many classifications marked the beginning of the classification approach. Finally, the speed and phase of renal activity were determined with the GFR testing approach. Chang et al. [23] developed two distinct ML models for predicting CKD with the prediction model and SVM. Based on a range of assessment metrics, it was determined that the SVM outperformed the other with an efficiency of 96.75 per cent.

Kennedy et al. [24] used various classification approaches to sort the data. They looked at how similar the traits were to create a connection matrix. Huang et al. [25] predicted renal disease with diverse learning approaches. They examined different algorithms to determine the most effective, and the research provides the outcome of MLP and C4.5 poses highest accuracy rates. Liu et al. [26] propose the combination of the (CKD) recognition components. Two important aspects, especially prognosis, enabled a more precise treatment approach by giving quantitative predictive risk factors. Naive Bayes, J48, and SOM were utilized in [27] to diagnose CKD early on. Using WEKA, a data analysis tool, experimentation with these systems on these resources was conducted. The finest algorithm for a trustworthy and early identification of CKD was determined solely to be J48. Of the three early-stage methods, it had the best accuracy.

Kong et al. [28] suggested a method for diagnosing chronic renal disease using the learning algorithm of the Adaboost ensembles. In this way of diagnosing, decision tree-based classification techniques were used. Mean absolute error, kappa, RMSE, and AUC were used to measure categorization effectiveness. The outcome shows that the Adaboost ensembles learning method produced superior classifier performance than human classifications. Six distinct fundamental classifications, including k-NN, NBs, SVM, RT, J48, DT and 3 different ensembles techniques including bagging, Ada Boost, and random substructure, were presented by Sherubha et al., [29]. The outcome showed that the randomized tree-based method and the J48-based technique with randomized subdomain & bagged ensemble learning both had 100% classifier accuracy. Chronic kidney disease (CKD) was forecasted by Sherubha et al. [30] using different classifiers such as the Artificial Neural Network (ANN) and Bayesian Network. Fast miners software was employed for experiments, and the findings indicated Naive Bayes produced more precise findings than the Convolutional Neural Network.

The system for encouraging decision-making was built and developed by Sherubha et al. [31], employing different classifiers such as probability-based NB,

BPN, LDA, decisions trees, lazy learners k-NN, and techniques for heuristic search categorization. On the Gathered the information, these algorithms' reliability was 78%, 81.5%, 90%, 93% and 76% correspondingly [32] – [35]. The effectiveness of their forecasting models still needs to be improved despite the previous work's valiant efforts to provide outcomes with a high degree of accuracy. Inability to forecast CKD with the best level of performance, a deep learning approach is anticipated in this study.

3. Methodology

This research methodology comprises essential steps like data acquisition, pre-processing, learning-based engineering, and categorization for evaluation. The design phases are provided in Fig 1.

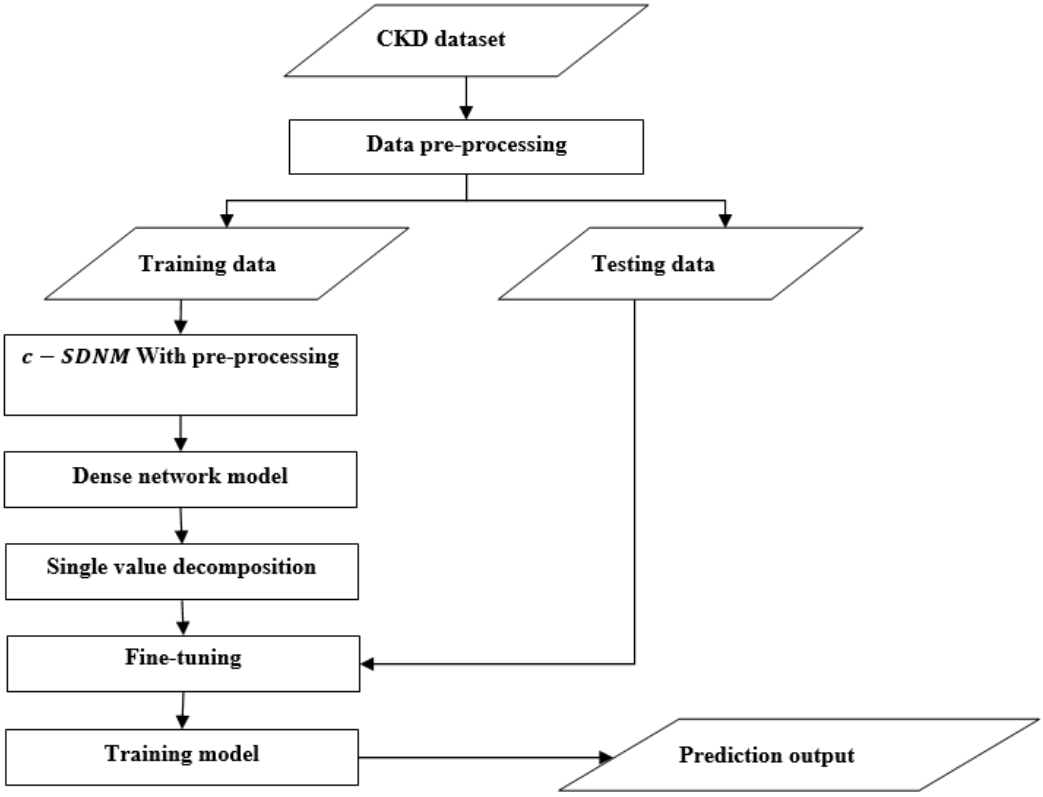

Fig Workflow of the proposed $c - SDNM$ model

3.1. Dataset

Utilizing UCI Dataset is the suggested model. It has 25 characteristics, 14 nominal, 11 numeric and 1 class attribute, representing 400 patients, 250 with CKD and 150 without. The utilized dataset is described in Table 1.

Table 1 Dataset description

ID	Attribute symbols	Description	Type
1	Age	Num	Yeats
2	Blood pressure	Num	Mm/Hg
3	Gravity	Nom	0.005 to 1.025
4	Albumin	Nom	0 – 5
5	Sugar	Nom	0 – 5
6	Red blood cells	Nom	Normal/abnormal
7	Pus cells	Nom	Normal/abnormal
8	Clumps	Nom	Yes/No
9	Bacteria	Nom	Yes/No
10	Blood glucose	Nom	Mg/dl
11	Urea	Num	Mg/dl
12	Creatinine	Num	Mg/dl
13	Sodium	Num	mEq/L
14	Potassium	Num	mEq/L
15	Haemoglobin	Num	gms
16	Cell volume	Num	gms
17	White blood cell (count)	Num	Cells/cum
18	Red blood cell (count)	Num	Millions/cum
19	Hypertension	Nom	Yes/No
20	Diabetes mellitus	Nom	Yes/No
21	Coronary artery disease	Nom	Yes/No
22	Appetite	Nom	Good or poor
23	Pedal edema	Nom	Yes/No
24	Anaemia	Nom	Yes/No
25	Class	Nom	CKD/Not

Num- Numerical and Nom- Nominal

3.2. Pre-processing

It might be difficult to categorize information that has incomplete data. The used dataset includes missing values, which leads to delays. Remove them before analyzing data. Missing values may be identified from cases or attributes. In cases(a record), discrepancies might be basic, medium, or complicated. If the case (a record) contains a maximum of one attribute with no values, it is a simple level. If the instance (a record) contains missing data for between 2% and 50% of the total characteristics, it is considered medium. While 50% to 80% of case (record) characteristics are missing, it's difficult. The discrepancies from the standpoint of the characteristics might be arbitrary, monotone, or univariate, if lacking data in one characteristic (feature). The pattern is monotone if there are gaps in at least three qualities. It is discretionary whether the blanks are in random properties [30]. The inference is seen as a means to get around problems with missing value situations. Missing value imputation keeps the whole sample size. Thus the dataset isn't diminished. The missing data may be imputed using various techniques with various attributes. Mode interpolation uses the most common value per column for conventional characteristics, while median imputation uses the mid-range value for statistical properties.

3.3. Customized network model

The proposed customized network model learns polynomial predictions using a deep network topology that approximates polynomial coefficients across training samples. Here is a quick overview of the $c-SDNM$technique. Fig2 depicts $c-SDNM$with four-layer structures.Let $\{(x, yi)\}_{i=1}^{m}$ be m samples of training, where x_i is a d −dimension and y_i.It corresponds to the label quantity. The definitions of the multivariate equations on R^d are:

$$p(x) = \sum_{i=1}^{\Delta} \sum_{\alpha^i} w_{\alpha^i} \prod_{l=1}^{d} x_l^{\alpha_l^i} \tag{1}$$

Where Δrepresents polynomial degree, i represents degree index, α^i represents d −dimensional index with $\sum_{l=1}^{d} \propto_l^{\ i} = i$,and w_{x^i} representsweighted index of α^i . In $c-PDNM$, we utilize $n_j^i(.)$ to express the j^{th} node's function as a component of its inputs in the i^{th} layer. The function generated in every node is either linear or weighted connection between two inputs to simplify. The set of values by degree-1 coefficients across training samples is provided when the first layer networks in$c-SDNM$ is constructed.

$$\{(< w, [1, x_1] >, \ldots., < w, [1, x_m] >); w \in R^{d+1} \tag{2}$$

Here,$(d+1)$ represents dimensional linear subspace of R^m. Here,$\langle\cdot,\cdot\rangle$ represents inner product. This linear independence set may be given a basis using the singular value decomposition (SVD) technique. We designate the matrix W that translates $[1\ X]$ into bias where 1 is vector and X is an array of observations, to optimize the model. The W column represent $d+1$ linear functions constituting the top layer of$c-SDNM$;for all $j = 1, \ldots, d+1$, the j^{th} node of the first layer function:

$$n_j^1(x) = < W_j, [1\ X] > \tag{3}$$

Here, $\left\{\left(n_j^1(x_1), \ldots, n_j^1(x_m)\right)\right\}_{j=1}^{d+1}$ specifies all the values produced by polynomials of degree 1 over training instances are based on a quantity of 1. Let $F_{i,j}^l = n_j^1(x_i)$ be the $m(d+1)$ matrix of F^l. Up to this point, a one-layer network has been constructed, with outputs that cover all values derived by linear models on the learning algorithm. Every degree-2 polynomial may be represented as:

$$\begin{aligned} &\sum_i \left(\sum_j \propto_i^{(g_i)} n_j^1(x)\right)\left(\sum_j \propto_s^{(h_i)} n_s^1(x)\right) + \left(\sum_j \propto_j^{(k)} n_j^1(x)\right) \\ &\qquad = \sum_{j,i} n_j^1(x) n_s^1(x) \left(\sum_i \propto_j^{(g_i)} \propto_s^{(h_i)}\right) + \sum_j n_j^1(x) (\propto_j^{(k)}) \end{aligned} \tag{4}$$

Where$\propto$represents scalar, and the superscripts g_i and h_i represent the polynomial degree with $g_i(x) = \sum_j \propto_i^{(g_i)} n_j^1(x)\ and\ h_i(x) = \sum_j \propto_s^{(h_i)} n_s^1(x)$. It specifies that the vector values are acquired from degree-2 polynomial encompasses the vector of

outcomes by the first layer node and productoutcomes of each pair of first-layer nodes. Additionally, we define $\tilde{F}^2$using Eq. (5):

$$\tilde{F}^2 = [(F_I^I \circ F_I^I) \dots (F_I^I \circ F_{|F_I|}^I) \dots (F_{|F_I|}^I \circ F_I^I) \dots (F_{|F_I|}^I \circ F_{|F_I|}^I)] \quad (5)$$

Where $\circ$ specifies the Hadamard product,$|F_I|$represents the number of columns in matrix F_I, and F_i represents the i^{th} column vector in matrix F, all representing the Hadamard product. The concatenation new matrix [FF2] is therefore inside the range of all values that degree-2 polynomials may achieve. Once again, SVD is used to create the foundation. Assume that F^2 is a subset of $\tilde{F}^2$ columns. F^2 creates the degree-2 polynomial basis, which may be derived using SVD to pick linear independent columns. The substitution - permutation network is then defined by the column of F^2. Each graph contains $F_i^I \circ F_j^I$, a second-layer junction that estimates the sum of first-layer nodes $n_i^1(.)$ and $n_j^1(.)$.Here, F is redefined as an enhanced matrix $[FF^t]$ for a straightforward notation. Then, for each t^{th} iteration, matrices F is kept, and its columns serve as the foundation for the values acquired by all degree-level functions$\leq (t-1)$. The new matrix is provided as:

$$\tilde{F}^t = [(F_1^{t-1} \circ F_1^1) \dots (F_1^{t-1} \circ F_{|F_1|}^1) \dots (F_{|F_1|}^{t-1} \circ F_1^1) \dots (F_{|F^{T-1}|}^{t-1} \circ F_1^1)] \quad (6)$$

Where,$[FF^t]$column form the$[F\tilde{F}^t]$ column. This recently created layer is added to build a network whose outcomes constitute the basis for the values produced by all degrees' functions ofdegree $\leq t$ over the training samples. $\tilde{F}^t$its then transformed to F^t via a matrix W of size $|F^{t-1}| \times |F^1|$ to maintain numerical stability as in Eq. (7):

$$F_s^t \coloneqq W_{i(s),j(s)} F_{i(s)}^{t-1} \circ F_{j(j)}^1 \; where \; s = 1,2,\dots,|F^t| \quad (7)$$

Where $i(s)$show that Wrepresents projection matrix that maps $\tilde{F}_t$into basis F_t of degree t polynomials. It is obtained via the Gram-Schmidt approach or reliable techniques, they also show that $'i'$ specifies$s$function and Wspecifies projection matrix. The construction of W in the first layer follows a similar process. A matrix F is complete after 1 repetition, and its columns serve as the foundation for all values obtained by degree polynomial$\leq (\Delta - 1)$ over the training samples. $c-SDNM$has a feed-forward neural design, which makes it fairly straightforward.

The learnt features from $n^{output}(+)$are input to a typical L_2-regularized hinge loss-oriented simple linear classification as the ultimate decision output. Naturally, additional classifiers may also be used to $c-SDNM$. Now, $c-SDNM$ is created using the approach above. The current technique, however, is confined to a tiny dataset with a smaller number of nodes. When $c-SDNM$ is applied to large training instances, the net width (number of nodes in every layer) grows, resulting in gigantic structures with significant computational overhead. To overcome this problem, an updated method employs a smaller fractional basis to produce tiny nodes' layer. In $c-SDNM$, it give upon precisely spanning $\tilde{F}_t$ and instead try to "approximately span" by employing the smaller incomplete basis with a finite size r, creating a layer with a width r. The network with sparse interconnections where nodes rely on two others, is computational quick. Consequently, this

innovative approach and the patchy node connectivity make $\tilde{F}_t$ especially useful for large-scale data while sacrificing efficiency and low computation cost for limited data advantages.

Algorithm 1:

1. Parameter initialization;
2. Evaluate singular value decomposition;
3. Predict W;

//End-to-end layer construction

4. for all $i \in \{1, \ldots |F^I|\}, n_i^1(x) = < W_b[1\ X] >$;
5. $F^1 = n_i^1(x)$; //independent linearly

//Perform stacked dense network construction

6. If error identified break

$$\tilde{F}^2 = [(F_I^I \circ F_I^I) \ldots (F_I^I \circ F_{|F_I|}^I) \ldots (F_{|F_I|}^I \circ F_I^I) \ldots (F_{|F_I|}^I \circ F_{|F_I|}^I)]$$

7. Construct layer by bias $F^t, W^t = (F, \tilde{F}_t)$
8. W^t is given as $F_s^t := W_{i(s),j(s)} F_{i(s)}^{t-1} \circ F_{j(j)}^1$
9. $F = [F\ F^t]$

//Output layer

10. Evaluate $w = \arg \min_{w \in R^{|F|^l}} (Fw, y)$;
11. $n^{output}(.) = < w, n(.) >$; $//n(.) = (n_1^1(.), \ldots, n_{|F^{t-1}|}^{t-1}(.))$
12. Predict error;
13. Perform validation;
14. **Return**$(n^{output}(.), error)$

When dealing with supervised $c - SDNM$, the first layer network computes the linear transformation that divides augmented training dataset $[1\ X]$ into $k$foremost source image using principal component analysis (PCA). The successive layer networks use the orthogonal least squares (OLS) method to choose the categories of $\tilde{F}_t$ most important for prediction. It uses a supervised methodology, where we iteratively choose the column of $\tilde{F}_t$ whose average, following projection onto the current basis F, has the strongest correlation with the lingering effects of the previous labels.

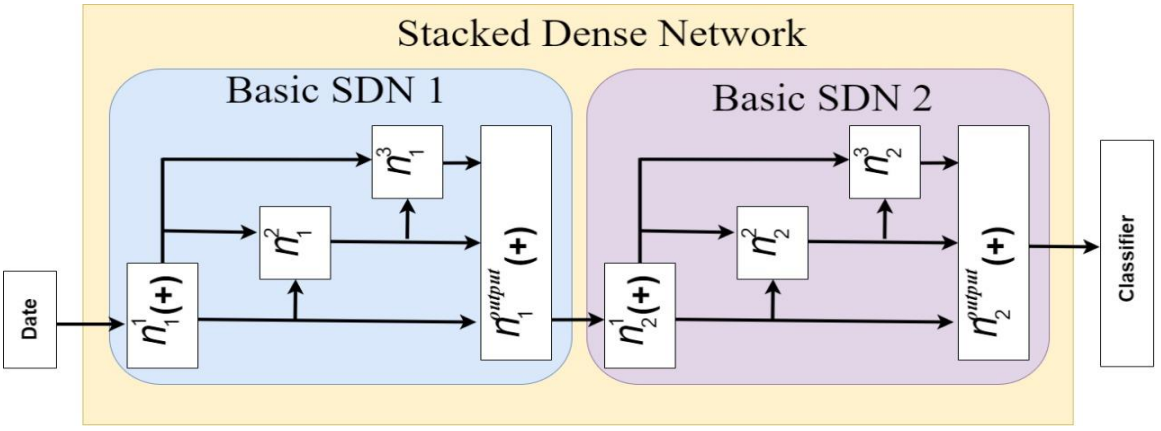

Fig 2 Stacked dense network model

3.4. Stacked dense network (SDN)

In $c-SDNM$, the description is enhanced by feeding the output of one layer into another layer, yielding a more complicated representation of data. This paper proposes $c-SDNM$ algorithm by borrowing DL's stacking approach. In $c-SDNM$, many basic DNMs may be stacked to create a deep hierarchy, with each DNM's output linked to the source of the next. A two-level $c-SDNM$ model is expanded to an m-level, as shown in Fig 2. For an m-level $c-SDNM$, the initial feature vector is fed into the first-level DNM block with n layers to produce the output of the learnt feature n_1^{output} (+), fed into the successive-level basic DNM. After the presenti^{th}-level basic DNM has finished training, output n_1^{output} (+), is given to the successive level of DNM for training. The fundamental DNM blocks, every composed of a straightforward and simple-to-learn component, are layered to create the total deep convolutional networks. Unlike previous deep designs, each basic DNM is learned block-wise without backpropagation. $c-SDNM$ has minimal computational cost compared to other DL back propagation techniques. It is important to note that the very first are there of each fundamental DNM still perform a transformation function splits augmented data matrix $[1\,X]$ into top k leader source image using PCA.To fuse and learn multisensory representations via $c-SDNM$, concatenate the vectors of various modalities. Nonetheless, this straightforward concatenation approach partially overlooks the variety of numerous modalities and cannot adequately investigate the complementary nature of different modalities or express their highly nonlinear relationships. So, as shown in Fig. 3, we suggest a two-stage $c-SDNM$-based DNM method.

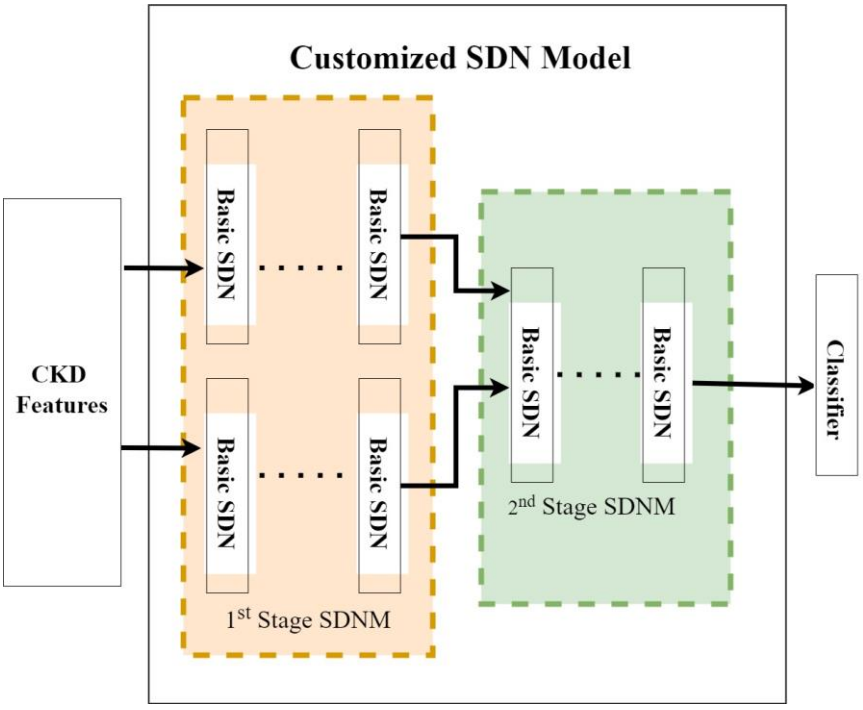

Fig 3 Customized SDN model

Every input data set will be given to the associated $c-SDNM$ module in the first step to learning high-level representations. Each unique modality's high-level properties represent its property without any correlational data. All learnt characteristics are then supplied to a new $c-SDNM$ module to associate all senses. Consequently, the inherent characteristics of each paradigm and the relationships between all modalities are included in the final learnt higher-level features. As a result, the characteristics that $c-SDNM$ learn are more robust and exclusionary.

This study uses a dataset often used in stacked CKD classification. Notably, the fusion technique in $c-SDNM$ differs from one in an SAE-based algorithm to find the disease. A pre-training approach is used with a percentage of malformed inputs. For instance, they conceal one modality by setting inputs to 0 and displaying the remainder of the training data with both models. The initial-level AE's hidden layer is trained to extract the original inputs that have been combined with hidden modalities. Original and distorted inputs are individually transmitted to higher network levels to ensure clean representation. Since DNM in our $c-SDNM$ method executes feed-forward supervised learning in every protocol stack devoid of fine-turning, it is challenging to implement the supervised learning to infer the associations among diseases. As a result, the concatenation features learnt in the first stage are concurrently trained with a second stage $c-SDNM$ to learn the social-democratic. It resembles the straightforward fusion technique used.

4. Numerical results and discussion

The dataset is splitrandomly into two parts, with the initial half containing 70% of the total information recorded for the model's training. The second half is for testing and comprises 30% of the dataset. Six performance indicators are utilized to verify the model. Accuracy, Recall, Precision and F-Measure are these measurements. Root Mean Square Error (RMSE) and Mean Absolute Error (MAE)are considered here. The mathematical model demonstrates how these measurements are calculated:

$$Accuracy = \frac{t_p + t_n}{t_p + f_p + f_n + t_n} \tag{8}$$

$$Precision = \frac{t_p}{t_p + f_p} \tag{9}$$

$$Recall = \frac{t_p}{t_p + f_n} \tag{10}$$

$$F - measure = \frac{2 * precision * recall}{precision + recall} \tag{11}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_t - y_t)^2}{n}} \tag{12}$$

$$MAE = \frac{|\hat{y}_t - y_t|}{n} \tag{13}$$

The terms true positive, false positive, false negative and true negative are $t_p, f_p, f_n, and\ t_n$.

Table 2 Performance evaluation

Metrics	Training	Testing
Accuracy	98.53	98.5
Precision	89.86	86.8
Recall	87.75	87.5
F-measure	98.5	98.6
RMSE	0.4850	0.492975
Mean Absolute Error	0.3765	0.380000

Table 3 Accuracy vs F1-measure comparison

Approaches	Accuracy	F1-measure
Naïve Bayes	95.5%	94%
J48	97.8%	97%
SMO	97.7%	96%
SVM	60.2%	97%
MLP	62.2%	95%
DT	63%	98%
Bayesian network model (BNM)	57%	93%
k-NN	58%	94%
ANN with GA	56%	93%
k-NN with GA	67%	97.5%
NN	98%	87%
SVM + GA	96%	95%
Gradient boosting	97.6%	96%
Deep belief network	98%	97%
$\boldsymbol{c-SDNM}$	**98.5%**	**98.6%**

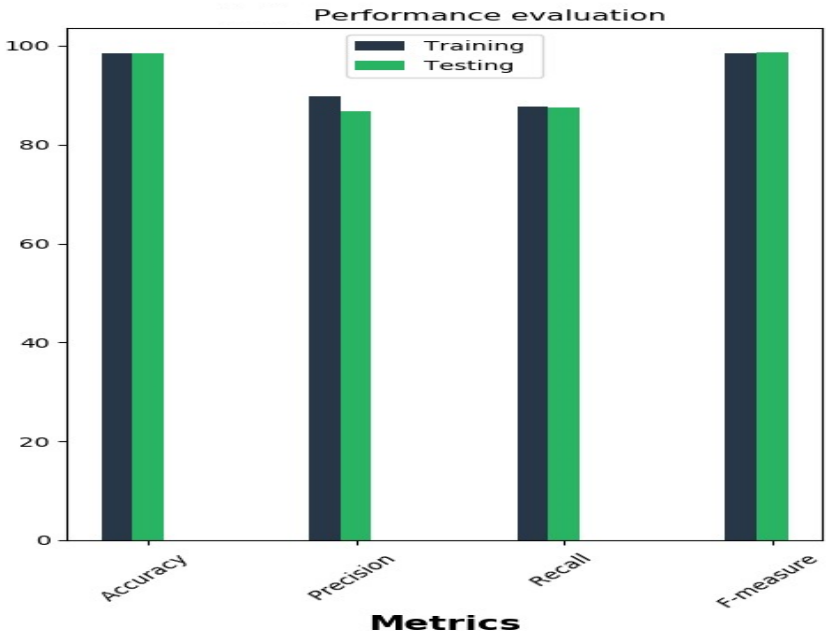

Fig 4 Performance evaluation

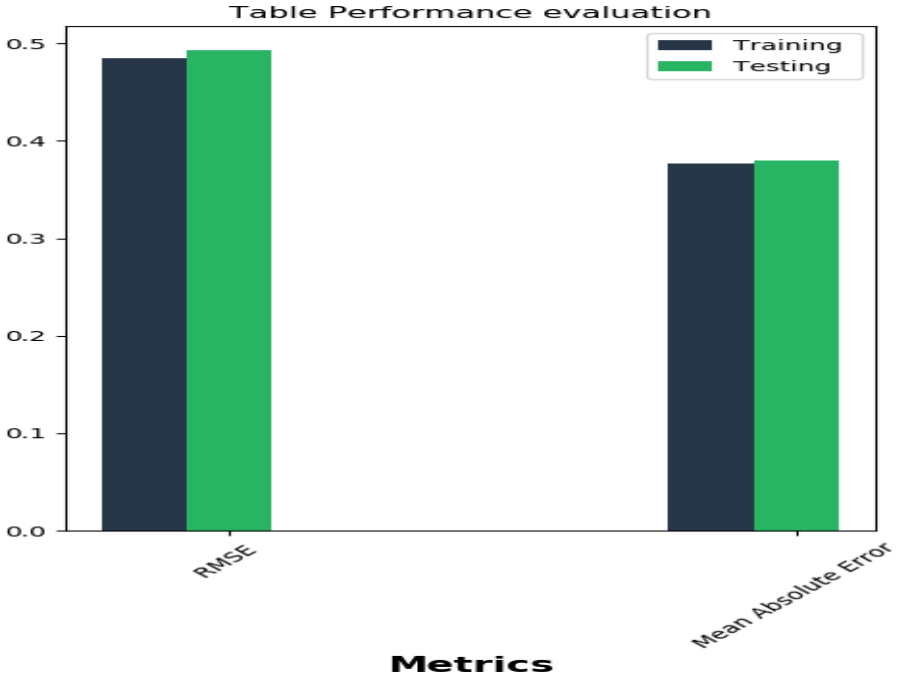

Fig 5 Performance evaluation for error prediction

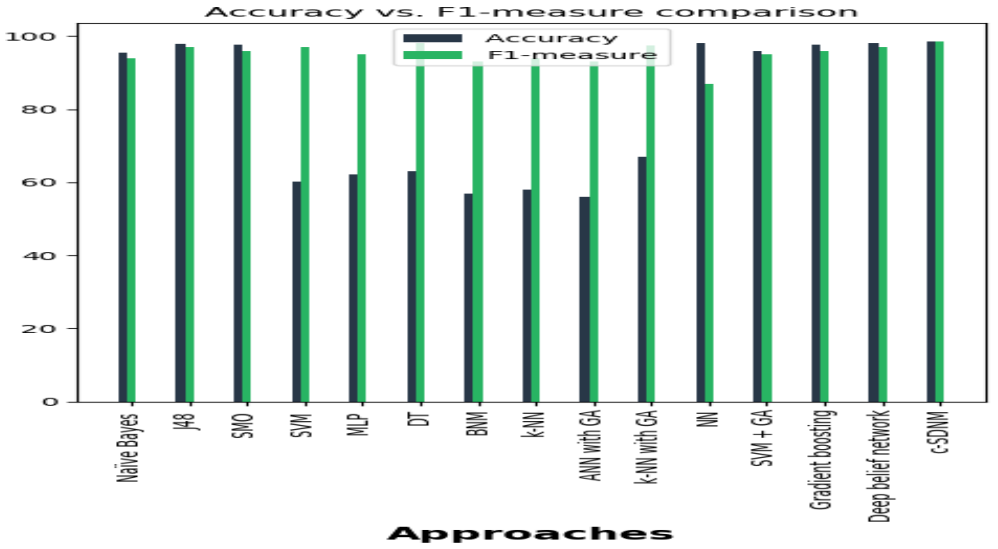

Fig 6 Accuracy and F1-measure comparison with other approaches

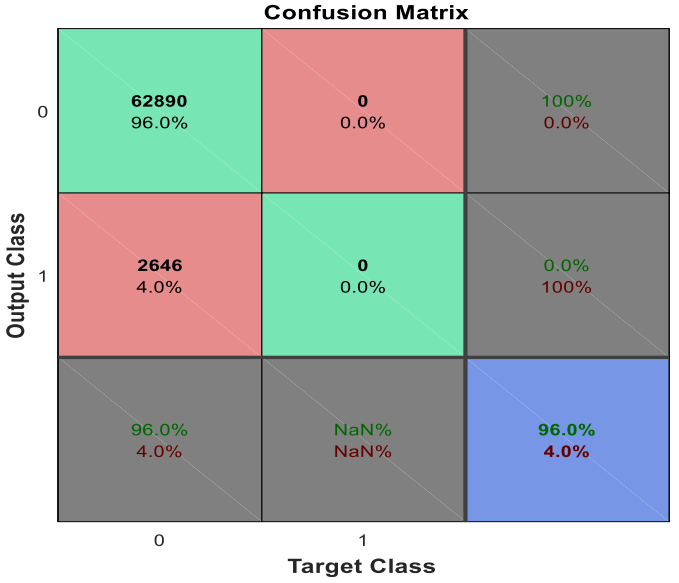

Fig 7 Confusion matrix

4.1. Analysis

We suggest the$c-SDNM$method in this study to efficiently train features for CKD diagnosis. Results from four sets of trials on the online database showed that the proposed $c-SDNM$method outperforms the most recent learning-based techniques (See Fig 4 to Fig 7).The initial ROI characteristics are low-level traits and cannot adequately discriminate between the characteristics of CKD. When using DNM to learn ROI characteristics, a more complicated representation is produced, leading to considerable gains. By stacking basic DNM blocks, higher-

level representations are generated. Therefore, for CKDdiagnosis$c-SDNM$ outperforms the original network model. Although,the 6-layer DNM is preferable to the 3-layer DNM with more hidden layers as the network increases computing cost without increasing approximation accuracy. The findings also indicate that $c-SDNM$ with two 3-layer DNMs outclasses the 6-layer DNMas the first layer of the second DNM is construct on the higher-level features, specifically the concatenation attributes of the firstDNM and it may produce more powerful and higher-level features after learning of the secondDNM. Additionally, $c-SDNM$ allows simpler parameter selection to attain the same effectiveness as DNM with development and improvement.

Since fusing multimodal neuroimaging data helps improve classification, a two-stage $c-SDNM$ approach is presented in Fig 4 to Fig 7. First, two $c-SDNM$s are applied to produce abstracts of every modality. To learn the merged characteristics, which comprise both the intrinsic characteristics of every modality and the connections between the attributes, all learnt characteristics then are focused on & given to a fresh $c-SDNM$. Compared with simply applying a single SDNM to the concentrated ROI characteristics of the dataset, the high-level characteristics acquired by the first DNMs will benefit and increase the cognitive capacity of the second $c-SDNM$ in DNM, and consequently, $c-SDNM$ can learn and integrate heterogeneous neuroimaging data more effectively.

In this work, $c-SDNM$performance can be evaluated using SVD and linear classifier classifiers. Both classifiers produce comparable results, indicating that AD classification relies upon learnt features more than classifiers. As a result, $c-SDNM$satisfactorily acquires a solid feature representation. Following are three key characteristics of DNM that lead to the great performance of $c-SDNM$. (1) The labelled observations in the neuroimaging collection are often tiny. DNM guarantees a well-trained deep network with minimal samples when applied to a dataset with few samples since the networks create small connections. (2) In DNM networks, the first k layers create a foundation for degree-k polynomials. The DNM network may thus pose a significant bias but may generally not overfit (i.e., have the lowest error), and even for improving service, the bias steadily decreases while the variance increases. As a result, in theory, overfitting might be reduced by managing the bias-variance balance. Furthermore, the intermediate layer's interconnections in DNM are highly shallow. Thus each node may only link to just a few other locations instead of all vertices inside the prior layer.DNM is suitable for tiny datasets as a result of its algorithmic nature. (3) As neuroimaging data generally give constraint labelled ground truth samples, and previous label knowledge is advantageous for classification with minimal data, supervised DNM is suited for small neuroscience datasets to unsupervised DL methods. The suggested $c-SDNM$ also outdoes the fine-turning-based supervised algorithms.

Here, DNM is a novel approach and information on its theory and advancement is currently limited. As a result, in our future work, we'll focus more on analyzingDNM's framework and, in particular, how it differs from other DL algorithms, rather than just continuing to enhance DNM as an algorithm. On the other hand, the $c-SDNM$ method suggested in this study demonstrates its efficacy for small datasets. In reality, DNM has established a solid track record for extracting features from massive amounts of data. Because there are no forward

or backward feedbacks between succeeding basic DNM, $c-SDNM$is quick and easy to use. They are anticipated to work well with vast amounts of data. We propose to use $c-SDNM$ to learn feature representation unswervingly from the dataset, similar to the technique. Semi-supervised $c-SDNM$ will also be researched in the future since it's simple to gather unlabeled medical input that enhances representation learning. We will attempt to integrate $c-SDNM$ for classification, prompted by the successful use of the input image. One may consider the learned representations from the second phase $c-SDNM$ and the separate learned elements from the dataset in the first stage $c-SDNM$s as multi-view data. As a result, theCKD characteristics may also be included, enhancing classification performance using $c-SDNM$ and MKL-based architecture.

5. Conclusion

This study uses a deep presumably make to create a calculated and the results of a neural network for predicting and classifying data on kidney disease. The model is created utilizing DNM and stacked network model. We leverage UCI's machine learning databases and pre-process missing data. A comparison with other models and an evaluation of the suggested$c-SDNM$ model's effectiveness are compared. With an accuracy of 98.52%, the suggested model outperforms the already available models. As a result, the suggested model offers an accurate prediction and classification for CKD. However, there is some research constraints related with the dataset acquisition. In future, the analysis can be done with two different datasets.

References

1. Ambekar and R. Phalnikar, “Disease risk prediction by using convolutional neural network,” in Proc. 4th Int. Conf. Comput. Commun. Control Autom. (ICCUBEA), Aug. 2018, pp. 1–5.
2. Carrio, C. Sampedro, A. Rodriguez-Ramos and P. Campoy, “A Review of Deep Learning Methods and Applications for Unmanned Aerial Vehicles”, Journal of Sensors, vol. 2017, pp. 1-13, 2017.
3. Chang and P. Lin, “Frail phenotype and mortality prediction: A systematic review and meta-analysis of prospective cohort studies”, International Journal of Nursing Studies, vol. 52, no. 8, pp. 1362-1374, 2015.
4. Diniari, N. K. S., & Aryani, L. N. A. (2022). Characteristics and pharmacological treatment options of delirium patients treated at Sanglah Central General Hospital. International Journal of Health & Medical Sciences, 5(1), 37-43. https://doi.org/10.21744/ijhms.v5n1.1835
5. Dixit, S. Srivastava and K. Chandra, “Prediction and Survival Rate Prostate Cancer Patient using Artificial Neural Network”, International Journal of Computer Applications, vol. 179, no. 5, pp. 19-21, 2017.
6. Elsayed and L. Syed, “An automatic early risk classification of hard coronary heart diseases using Framingham scoring model,” in Proc. 2nd Int. Conf. Internet Things, Data Cloud Comput., Mar. 2017, pp. 1–8.
7. Frohlich and P. J. Quinlan, “Coronary heart disease risk factors: Public impact of initial and later-announced risks,” Ochsner J., vol. 14, no. 4, p. 532, 2014

8. Gonsalves, F. Thabtah, R. M. A. Mohammad, and G. Singh, “Prediction of coronary heart disease using machine learning: An experimental analysis,” in Proc. 3rd Int. Conf. Deep Learn. Technol., 2019, pp. 51–56.
9. Huang, L. Shi and J.A.K. Suykens, “Sequential minimal optimization for SVM with pinball loss”, Neurocomputing, vol. 149, pp. 1596-1603, 2015.
10. Huang, W. Dong, H. Duan and J. Liu, “A Regularized Deep Learning Approach for Clinical Risk Prediction of Acute Coronary Syndrome Using Electronic Health Records”, IEEE Transactions on Biomedical Engineering, vol. 65, no. 5, pp. 956-968, 2018.
11. Huang, Y. Wei, J. Yi, and M. Liu, “An improved kNN based on class contribution and feature weighting,” in Proc. 10th Int. Conf. Measuring Technol. Mechatronics Autom. (ICMTMA), Changsha, China, Feb. 2018, pp. 313–316.
12. Huang, Z. Ge, W. Dong, K. He, H. Duan and P. Bath, “Relational regularized risk prediction of acute coronary syndrome using electronic health records”, Information Sciences, vol. 465, pp. 118-129, 2018.
13. Huang, Z. Ge, W. Dong, K. He, H. Duan and P. Bath, “Relational regularized risk prediction of acute coronary syndrome using electronic health records”, Information Sciences, vol. 465, pp. 118-129, 2018.
14. Huling, M. Yu, Menggang, M. Liang and M. Smith, “Risk prediction for heterogeneous populations with application to hospital admission prediction: Risk Prediction for Heterogeneous Populations”, Biometrics, vol. 74, no. 2, pp. 557-565, 2018.
15. Kennedy, W.L. Wiitala, R.A. Hayward, J.B. Sussman, “Improved Cardiovascular Risk Prediction Using Nonparametric Regression and Electronic Health Record Data”, Medical Care, vol. 51, no. 3, pp. 251- 258, 2013
16. Khandoker, Y. Al Zaabi, and H. Jelinek, "What can tone and entropy tell us about the risk of cardiovascular diseases?" Proc. Comput. Cardiology Conf. (CinC), Dec. 2019, pp. 1–4.
17. Kim, YG Lee, JW. Kim, JJ. Park, B. Ryu and JW. Ha, “Highrisk Prediction from Electronic Medical Records via Deep Attention Networks”, in 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017
18. Kong, J. Gao, Y. Xu, Y. Pan, J. Wang and J. Liu, “Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier”, Neurocomputing, vol. 324, no. 9, pp. 63-68, 2019.
19. Krishnani, A. Kumari, A. Dewangan, A. Singh, and N. S. Naik, “Prediction of coronary heart disease using supervised machine learning algorithms,” in Proc. IEEE Region 10 Conf. (TENCON), Oct. 2019, pp. 367–372.
20. Lee, S. Jung, M. Kim, and S. Kim, "Synthetic minority over-sampling technique based on fuzzy c-means clustering for imbalanced data," in Proc. Int. Conf. Fuzzy Theory Appl. (fuzzy), Nov. 2017, pp. 1–6
21. Li, J. Li, X. Lan, Y. An, W. Gao and Y. Jiang, “Experiences of building a medical data acquisition system based on two-level modeling”, International Journal of Medical Informatics, vol. 112, pp. 114-122, 2018.
22. Liu, F. Wu and J. Wang, “Efficient Multi-Kernel DCNN with Pixel Dropout for Stroke MRI Segmentation”, Neurocomputing, 2019.

23. Martin-Isla, V. M. Campello, C. Izquierdo, Z. Raisi-Estabragh, B. Baeßler, S. E. Petersen, and K. Lekadir, “Image-based cardiac diagnosis with machine learning: A review,” Frontiers Cardiovascular Med., vol. 7, p. 1, Jan. 2020.
24. Quesada, A. Lopez-Pineda, V. F. Gil-Guillén, R. Durazo-Arvizu, D. Orozco-Beltrán, A. López-Domenech, and C. Carratalá-Munuera, “Machine learning to predict cardiovascular risk,” Int. J. Clin. Pract., vol. 73, no. 10, Oct. 2019, Art. no. e13389.
25. Quesada, A. Lopez-Pineda, V. F. Gil-Guillén, R. Durazo-Arvizu, D. Orozco-Beltrán, A. López-Domenech, and C. Carratalá-Munuera, “Machine learning to predict cardiovascular risk,” Int. J. Clin. Pract., vol. 73, no. 10, Oct. 2019, Art. no. e13389.
26. Reddy. D, D. B. Raveendrababu, and D. A. Govardhan, “A model for improving classifier accuracy using outlier analysis,” Int. J. Comput. Technol., vol. 7, no. 1, pp. 500–509, May 2013.
27. Shadmi, V. Mazo, O. Bregman-Amitai, and E. Elnekave, ''Fullyconvolutional deep-learning-based system for coronary calcium score prediction from non-contrast chest CT,'' in Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI), Apr. 2018, pp. 24–28.
28. Sharma, S. Verma, and Kavita, ''Prediction of heart disease using Cleveland dataset: A machine learning approach,'' Int. J. Recent Res. Aspects, vol. 4, no. 3, pp. 17–21, 2017.
29. Sherubha, “An Efficient Intrusion Detection and Authentication Mechanism for Detecting Clone Attack in Wireless Sensor Networks”, Journal of Advanced Research in Dynamical and Control Systems (JARDCS), Volume 11, issue 5, Pg No. 55-68
30. Sherubha, “An Efficient Network Threat Detection and Classification Method using ANP-MVPS Algorithm in Wireless Sensor Networks”, International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-8 Issue-11, September 2019
31. Sherubha, “Graph Based Event Measurement for Analyzing Distributed Anomalies in Sensor Networks”, Sådhanå(Springer), 45:212, https://doi.org/10.1007/s12046-020-01451-w
32. Song, J. Warren, and P. Riddle, ''Developing high-risk clusters for chronic disease events with classification association rule mining,'' in Proc. 7th Australas. Workshop Health Inform. Knowl. Manage., vol. 153, 2014, pp. 69–78.
33. Sultana, A. Haider, and M. S. Uddin, “Analysis of data mining techniques for heart disease prediction,” in Proc. 3rd Int. Conf. Electr. Eng. Inf. Commun. Technol. (ICEEICT), Sep. 2016, pp. 1–5.
34. Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2021). The COVID-19 pandemic. International Journal of Health Sciences, 5(2), vi-ix. https://doi.org/10.53730/ijhs.v5n2.2937
35. Vasiljeva and O. Arandjelovic, “Diagnosis Prediction from Electronic Health Records Using the Binary Diagnosis History Vector Representation”, Journal of Computational Biology, vol. 24, no. 8, pp. 767-786, 2017.
36. Xu, Y. Zhang, M. Azhar, P. ZHANG, Y. Li and S. Khatoon, “Data Mining on ICU Mortality Prediction Using Early Temporal Data: A Survey”, International Journal of Information Technology Decision Making, vol. 16, no. 1, pp. 117-159, 2017.