Received: July 4, 2022. Revised: August 21, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Deep Ensemble Learning with GAN-based Semi-Supervised Training Algorithm for Medical Decision Support System in Healthcare Applications

Periya Nayaki Anbarasu[1]* **Thanabal Muniyandi Suruli[1]**

[1]*Department of Computer Science and Engineering,*
PSNA College of Engineering and Technology, Dindigul, Tamilnadu, India.
*Corresponding author's Email: periyanayaki.tdu.phd@gmail.com

Abstract: Nowadays, clinical decision support system (CDSS) is emerged broadly due to the growth of smart applications. Yet, labeling a massive amount of clinical data is expensive and needs domain expertise. Also, how to design a robust CDSS model in a healthcare platform with a minimal number of annotated clinical data was an essential problem. To tackle this issue, a generative adversarial network (GAN)-based semi-supervised training model was developed, which increases the quantity of annotated data and solves the unbalanced annotated labels. As well, support vector machine (SVM) and K-nearest neighbor (KNN) classifiers were applied as the base learners to predict the label for unannotated data. But, the result from both base learners was not often similar, which impacts the prediction efficiency. So, this article introduces a deep ensemble learner model (DELM) for predicting labels to the unannotated data by stacking multiple machine learning classifiers as base learners. In this model, two levels of prediction: the initial level and the second level. During the initial level, multiple base learners including SVM, KNN, naive bayes (NB) and random forest (RF) learn the dataset with both the annotated and unannotated samples independently to predict the class of the unannotated instances. Then, those prediction outcomes are merged during the second level, which applies the deep neural network (DNN) as an ensemble classifier to get the final predicted class of the unannotated instance. Finally, the experimental results exhibit that the DELM with GAN achieves 86.54 %, 84.83 % and 86.72 % accuracies on SPECT, WDBC and Hallmarks databases, correspondingly, compared to the Fuzzy-AHP+ANN, HTM+LSTM, DBSCAN+SMOTE-ENN+XGBoost and GAN-based semi-supervised models.

Keywords: Clinical decision support, Deep learning, Ensemble, GAN, Label prediction, Machine learning, Semi-supervised training.

1. Introduction

Several successful clinical initiatives have been created in the last decade. These kinds of initiatives make healthcare more accessible [1]. People will also use information and communication technologies to make it easier for patients and their clinicians to communicate, boosting the patient's chances of survival. Doctors may easily access patients' medical histories, laboratory tests, pictures, and medication information from anywhere and at any time. Patients may continue to provide information regarding their safety and exposure to the medical condition. The CDSS is an interactive, flexible, and adaptable computer-based information system [2]. CDSSs are designed to assist physicians and other healthcare workers in making diagnoses and making decisions. The key advantage of this approach is that it enhances patient care and healthcare professional efficiency [3].

The CDSS is an interactive technique that generates health-related recommendations based on patient information variables. The basic purpose of advanced medical decision support systems is to assist doctors in making clinical decisions. This means that experts collaborate with a CDSS to examine patient data and make a diagnosis. CDSSs were once thought to be used to assist physicians in

making decisions [4]. The new approach of enlisting their assistance indicates that the doctor collaborates with the CDSS to perform a complete study of the patient's details that either a person or the CDSS might achieve on their individual.

The CDSS is an example of how a clinician might treat a CDSS. The CDSS examines part of the patient's information before presenting a list of possible diagnoses. The physician evaluates which diagnoses are relevant and which are not using the CDSS output and, if necessary, conducts additional testing. Pre-diagnosis CDSS systems in medicine are used to help doctors make diagnoses. During diagnosis, the CDSS assists clinicians in examining and filtering their early diagnostic choices to improve their outcomes [5].

Briefly, the main role of the CDSS is the categorization process in which a categorizer trained on the annotated data (e.g., patient's record) with many attributes (e.g., diagnosis outcomes) is implemented to allocate the tag to an uncertain data (e.g., new patients) and the tag computes one of the deliberate labels defining probable treatments. Numerous CDSSs [6-10] were designed and utilized for the diagnosis of multiple diseases, particularly tumors. Generally, better categorization was established from more available data defining prior experience. But, many realistic problems related to the clinical data gathered from the healthcare systems were addressed: (i) the quantity of clinical data gathered by healthcare systems is very huge, yet the quantity of annotated set was relatively less; and (ii) the clinical database acquired in the medical system was rather varied, which often tends to the class imbalance.

On the small quantity of annotated samples and a huge quantity of unannotated samples, semi-supervised training [11] was an appropriate solution for developing a robust classifier. But, it usually operates under specific hypotheses and a minimum ratio of the annotated set was needed to estimate such hypotheses to achieve the semi-supervised training process. When faced with an unbalanced dataset, most of the training methods focus on the majority class while neglecting the importance of the minority class. Oversampling or under-sampling methods were utilized in a few sampling-based procedures to balance classes [12-13]. However, which sampling approach was best for the target database remains unsolved because of their limitations. The under-sampling neglects specific relevant samples and the oversampling was prone to overfit. In contrast, the SMOTE was a common synthetic data creation technique, which creates a random quantity of artificial minority samples to skew the categorizer's training bias toward minority labels [14]. But, it aims to generate the synthetic samples by simply considering the vicinity of minority labels and the label populace, while ignoring the allocation of samples from the majority labels. Also, the key challenges include: (i) the creation of plenty of samples around unwanted minority samples may not aid the minority class training; (ii) the noisy samples may be involved in regions of the majority classes and (iii) the increasing overlapping among classes disrupts the class margins. To combat these challenges, Yang et al. [15] developed the GAN-based semi-supervised learning model to support healthcare solutions. The GAN was utilized to augment the quantity of tagged information and solve the imbalanced tagged classes with extra counterfeit information to enhance the semi-supervised learning efficiency. Also, the efficiency of estimated tags on the untagged dataset was increased by merging the strengths of co-learning and self-learning in a semi-supervised learning model. On the other hand, linear SVM, radial basis function SVM (RBF-SVM) and KNN algorithms were utilized as base learners during label prediction. According to the fact that each of these algorithms may outperform others or include limitations in distinct scenarios, it is expected that an ensemble model that takes benefits of multiple learners will provide superior efficiency.

Hence in this paper, a novel CDSS is developed, which introduces deep learning to an ensemble learner model to stack multiple distinct machine learning classifiers to enhance the efficiency of predicting the label for unannotated data. This model consists of 2 major levels: (1) the initial level prediction models and (2) the second level prediction models. In the initial-level prediction model, the dataset with both the annotated and unannotated instances is provided to multiple base learners such as SVM, KNN, NB and RF, each of which provides the predicted class of the instances. In the second-level prediction model, the DNN classifier is utilized as an ensemble model rather than a weighted average or majority voting scheme to fuse the predictions in the initial level to minimize the generalization error and obtain a more accurate class label for unknown data. Thus, this model can automatically learn both linear and nonlinear correlations among the instances to predict the class label efficiently.

The residual portions of this manuscript are arranged as the following: Section 2 reviews works related to the CDSS using machine and deep learning algorithms. Section 3 describes the presented model and section 4 exhibits its

effectiveness. Section 5 concludes this study and provides future improvements.

2. Literature survey

A CDSS method was designed [16] to diagnose cardiac arrhythmia using discrete wavelet transform (DWT) and random forest (RF) classification. First, the ECG signals were decomposed into multiple spectrum bands by the DWT. Then, the frequency bands were acquired to represent the distribution of wavelet coefficients using the various statistical attributes. Further, the RF classifier was applied to classify ECG signals to identify the heart rhythm. But, it has a high computation burden while using large-scale databases.

An integrated CDSS model was designed [17], which measures various heart failure traits by an expert cardiac physician. First, a fuzzy analytic hierarchy process (Fuzzy_AHP) was employed to compute the global weights for the traits according to their separate contribution. Then, such global weights were utilized to train the artificial neural network (ANN) model for predicting the heart failure risks in patients. But, it was not able to optimize the hidden neurons automatically and also not suitable for a vast amount of instances.

An IoT with cloud-based CDSS using deep neural network (DNN) model was presented [18] to predict chronic kidney disorder with its severity range. Also, a particle swarm optimization (PSO)-based attribute choice method was employed to optimize the DNN variables and enhance their efficiency. But, its complexity was high when considering more instances.

CDSS using 3D deep convolutional neural network (DCNN)-based was designed [19] to recognize the lung nodule. In this model, a new median intensity projection was applied to leverage 3D data from CT scans and an advanced multi-region proposal network was adopted to select the promising region-of-interest (ROI). Also, a cloud paradigm was included to enhance the robustness. But its accuracy was relatively not effective to detect micro-nodules.

A diagnostic fog-based IoT paradigm was implemented [20] to monitor type-2 diabetes patients. As well, a new CDSS was introduced to predict type-2 neuromorphic values using the VIKOR scheme. On the other hand, it needs to design deep learning models to enhance accuracy.

CDSS for real-time monitoring of mean arterial pressure (MAP) status was designed [21] at the bedside using a novel machine-learning algorithm. Initially, online learning was applied by hierarchical temporal memory (HTM) to enable real-time stream processing and obtain unsupervised predictions. Then, the long short-term memory (LSTM) classifier was applied to monitor the status of the patient's MAP according to the online predictions. But, it was prone to overfit and needs more memory to train.

A powerful heart disorder estimation framework was designed [22] that involves density-based spatial clustering of applications with noise (DBSCAN) to identify and remove the artifacts, a hybridized synthetic minority over-sampling technique-edited nearest neighbor (SMOTE-ENN) to compensate the learning samples and XGBoost to recognize heart disorder. But, it was limited to the particular databases, which influences the model efficiency.

A mobile application-based CDSS was [23] developed, which was leveraged on the deep learner to assist zone providers to estimate bedsores, categorize their condition, forecast their progress together the period and create proper solutions regarding the group of activities to efficiently diagnose them. But, the training data was inadequate, which impacts the accuracy.

A new CDSS was developed [24] depending on deep learning and neural network schemes for the diagnosis of temporomandibular joint (TMJ) disorder. A non-invasive system was adopted for the recording of TMJ sounds. An interface was designed that will enable the dentist to work on the recorded audio data. Then, signal processing, ANN and deep learning models were utilized to categorize TMJ data. But, its success rate was still not efficient due to the limited amount of data.

A light gradient boosting machine (LightGBM)-based CDSS was designed [25], which reduces the patient risk and cost to the medicinal system by enhancing the diagnostic field of invasive coronary angiography via an optimized outpatient choice. But, it needs external validation before being executed clinically. Also, it did not apply in other medical areas without a few local adaptations, because this model should learn doctor referral patterns from local data

2.1 Research contribution

This research concentrates on improving the efficiency of the CDSS for predicting the appropriate diseases and recommending proper diagnosis. This is achieved by introducing the semi-supervised learning-based data augmentation model, which generates more annotated data to solve the class imbalance problem and train the ensemble

Table 1. Notations used in this study

Symbols	Description
G	Generator
D	Discriminator
P_G	Generator distribution
z	Noise parameter
x	Target sample space
n_l	Number of units in the DNN
L_l	Unit l in the DNN
L_1	Input unit in the DNN
L_{n_l}	Output unit in the DNN
s_l	Number of neurons in unit l
W, B	Variables of the DNN
W_{ij}^l	Weight related to the link between neuron j in unit $l-1$ and neuron i in l
B_i^l	Bias of unit i in l
$(x^1, y^1), \ldots, (x^m, y^m)$	Training collection
m	Number of instances
$J(W, B)$	Objective function of the DNN
λ	Weight decay variable
$h_{W,B}(x)$	Nonlinear hypothesis
$f: \mathbb{R} \to \mathbb{R}$	Activation factor
a_i^l	Activation of unit i in l
z_i^l	Weighted sum unit i in l
x_i	Neuron i in the input unit L_1
$a_i^{n_l}$	Activation of the neuron in the output unit
α	Learning rate
$\delta_i^{n_l}$	Residual for all neurons i in n_l
δ_i^l	Residual for all neurons i in l
$\frac{\partial J(W,B;x,y)}{\partial W_{ij}^l}, \frac{\partial J(W,B;x,y)}{\partial B_i^l}$	Target partial derivative
D	Entire database
D_i	Sub-collections of database D
$h_i(x_i)$	Binary factor of i^{th} model
D_i'	Fresh database
C	Final classifier

classifier efficiently for disease prediction. The scientific contribution of this research is the following:

- Initially, the medical data is fed to the GAN for creating the synthetic annotated instances, which solves the class imbalance problem in the raw datasets.
- Then, the obtained data instances are learned by the DELM-based classifier for predicting and diagnosing multiple types of diseases.
- This ensemble classifier model can handle large-scale datasets and solve the class-imbalanced problems efficiently rather than each classifier alone for the CDSS. As a result, the accuracy of classifying the accurate diseases from the large-scale dataset is increased.

The below section briefly explains this contribution to the CDSS. Table 1 presents the notations used in this study.

3. Proposed methodology

In this section, the DELM with GAN-based semi-supervised training algorithm is explained briefly according to its execution information on limited annotated clinical records and imbalanced labels. Then, the structure of CDSS depending on this presented DELM-GAN-based semi-supervised algorithm is presented.

3.1 Semi-supervised training algorithm based on DELM-GAN

In this presented algorithm, DELM-based base learner trained on the annotated collection is initially introduced to categorize the unannotated collection. Later, the true unannotated sample having similar pseudo tags that are allocated by the DELM-based support learner, are included into the annotated collection for consecutive iterations of training. If the annotated collection has not updated after a number of iterations, the unannotated collection will be neglected when it is not empty. This termination process can considerably avoid efficiency loss experienced using the unannotated collection having indecisive pseudo tags.

Typically, it is rather difficult to learn the semi-supervised training algorithms on unbalanced and very tiny part of annotated collection. So, this difficulty is resolved by applying GAN, which dynamically creates the training collection, wherein DELM-based base learner is effectively trained on, by including synthetic data to the annotated collection at all iterations. Once the annotated collection is augmented, GAN is utilized another time to fill the augmented annotated collection and balance its minority labels. Moreover, an absolute categorizer is acquired from the well-arranged annotated collection.

3.1.1. Generative adversarial network

Normally, GAN comprises generator (G) and discriminator (D) models. G creates synthetic instances that are acquired from an equal distribution as the learning data, whereas D evaluates whether an instance comes from actual data or synthetically generated by G.The fitness function of GAN is defined by

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim P_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim P_z(z)}\left[1 - log\, D\left(G(z)\right)\right] \quad (1)$$

In Eq. (1), G intends to create instances $G(z)$ from their distribution P_G and it converts the input noise parameter z into the target sample space x. This G is learned via gaming against D that intends to differentiate between instances from the accurate sample allocation P_{data} and P_G. Thus, for a considered G, the optimal D is $D(x) = P_{data}(x)/P_{data}(x)P_G(x)$, whereas the optimal G is achieved if $P_G(x) = P_{data}(x)$.

3.1.2. Deep ensemble learning model as base learner

This DELM consists of 2 levels. In the first level, 4 well-known machine learning classifiers are applied as the initial-level prediction models. Such classifiers are SVM, KNN, NB and RF.

(i). KNN is a non-parametric prediction technique, which is applied if there is litter or no prior information about the data distribution. It converts instances to a metric space in which gaps among instances are computed. The gap factor involving a test instance and the learning instances is the support through determining which, it categorizes the test instance depending on the most familiar label in its k-nearby learning instances.

(ii). SVMs transfer the input vector into a high-dimensional attribute space and find a hyperplane, which divides the instances into 2 groups. The distance involving 2 groups is as large as probable. After that, new instances are transformed into an equal space and expected to fit to the class depending on which plane of the distance they drop on with high belief.

(iii). NB is used to build classifiers or predictors. It allocates labels to unknown instances, defined as vectors of feature values, while the labels are acquired from few finite sets.

(iv). RFs are an ensemble training algorithm, which merges tree estimators, all of which depend on the ranges of an arbitrary vector sampled separately and with an equal distribution. The absolute result is the most familiar label, which accepts the majority of votes from the trees in the forest, so providing an effective system.

In the second level, these multiple classifiers are stacked to learn the predictions from the initial level optimally and obtain the absolute prediction. To achieve this, the DNN is adopted as the ensemble classifier, which stacks many categorizers. This

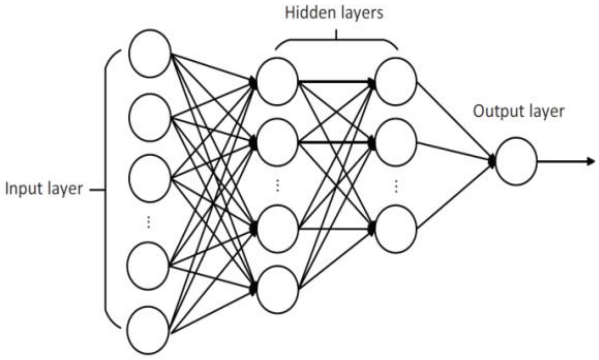

Figure. 1 Typical DNN structure

DNN is trained to create a result as a mixture amid the input attributes. For a collection of attributes and a label, it learns to be a nonlinear function approximators, wherein specific or multiple nonlinear units known as hidden units are implemented between the input and output units. The typical DNN structure is portrayed in Fig. 1.

The leftmost unit is the input unit having neurons known as input neurons. The rightmost unit is the output unit having a result neuron. The mid units are hidden units, which are created by the secret neurons. To categorize instances properly, an objective factor is defined that determines the loss between the expected and real values. After that, via learning by the learning instances, the system changes the values of its inner adaptable variables, which represent the input-output factor to minimize the loss.

In this process, the stochastic gradient descent (SGD) strategy is utilized. In the DNN, the quantity of units is represented as n_l and unit l as L_l, therefore unit L_1 is the input unit and unit L_{n_l} is the output unit. Consider s_l is the quantity of neurons in unit l. The DNN consists of variables $W = \{W^1, \dots, W^{n_l}\}$ and $B = \{B^1, \dots, B^{n_l}\}$, where $W_{ij}^l, j = 1, \dots, s_{l-1}, i = 1, \dots, s_l, l = 2,3, \dots, n_l$ is the weight related to the link between neuron j in unit $l-1$ and neuron i in l, as well as, $B_i^l, i = 1, \dots, s_l, l = 2,3, \dots, n_l$ is the bias of unit i in l.

Let training collection $\{(x^1, y^1), \dots, (x^m, y^m)\}$ of m instances, with which the DNN is trained by the SGD. The objective function is described by

$$J(W,B) = \frac{1}{m}\sum_{i=1}^{m} J\left(W,B;x^i,y^i\right) + \frac{\lambda}{2}\sum_{l=2}^{n_l}\sum_{j=1}^{s_{l-1}}\sum_{i=1}^{s_l}\left(W_{ij}^l\right)^2 = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\left\|h_{W,B}\left(x^i\right) - y^i\right\|^2\right) +$$

$$\frac{\lambda}{2}\sum_{l=2}^{n_l}\sum_{j=1}^{s_{l-1}}\sum_{i=1}^{s_l}\left(W_{ij}^{l}\right)^2 \quad (2)$$

In Eq. (2), the initial expression is the mean square error (MSE) and the 2nd expression is the normalization utilized to restrict the magnitudes of the weights and avoid overfitting, as well as, λ denotes the weight decay variable that controls the associative significance of these 2 expressions. The nonlinear hypothesis $h_{W,B}(x)$ of the DNN is represented by

$$h_{W,B}(x) = f(W^T x + B) \quad (3)$$

In Eq. (3), $f: \mathbb{R} \to \mathbb{R}$ is known as the activation factor. In this model, the rectified linear unit (ReLU) is utilized, i.e. $f(z) = max\{0, z\}$. For case, the activation of unit i in l is defined as a_i^l and the weighted sum is represented as z_i^l, so

$$a_i^l = f\left(z_i^l\right) = f\left(W_{i1}^{l-1}a_1^{l-1} + W_{i2}^{l-1}a_2^{l-1} + \cdots + W_{is_{l-1}}^{l-1}a_{s_{l-1}}^{l-1} + B_i^{l-1}\right) \quad (4)$$

Also, using x_i as the neuron i in the input unit L_1, i.e.$a_i^1 = x_i$. So, the activation of the neuron in the output unit is as:

$$h_{W,B}(x) = a_i^{n_l} = f\left(W_{i1}^{n_l-1}a_1^{n_l-1} + W_{i2}^{n_l-1}a_2^{n_l-1} + \cdots + W_{is_{n_l-1}}^{n_l-1}a_{s_{n_l-1}}^{n_l-1} + B_i^{n_l-1}\right) \quad (5)$$

This stage to determine the activation of all units is known as the forward propagation. In SGD, the aim is to reduce $J(W, B)$ by changing variables W and B. Initially, all W_{ij}^l and B_i^l are initialized to a tiny arbitrary range about 0 and modify the variables in all iterations of SGD as:

$$W_{ij}^l = W_{ij}^l - \alpha\frac{\partial J(W,B)}{\partial W_{ij}^l} \quad (6)$$

$$B_i^l = B_i^l - \alpha\frac{\partial J(W,B)}{\partial B_i^l} \quad (7)$$

In Eqs. (6) and (7), α denotes the learning rate. After that, the back-propagation scheme is applied to determine the partial derivatives. For a learning data (x, y), the back-propagation scheme is explained below.

- Perform the forward propagation computations to determine the activation of all neurons in L_2 to the output unit L_{n_l}.
- For all neurons i in n_l, the residual is determined as:

$$\delta_i^{n_l} = \frac{\partial}{\partial z_i^{n_l}}\frac{1}{2}\left\|y - h_{W,B}(x)\right\|^2 = -\left(y_i - a_i^{n_l}\right)f'\left(z_i^{n_l}\right) \quad (8)$$

- For all neurons i in $l, l = n_l - 1, n_l - 2, \ldots, 2$, the residual is determined as:

$$\delta_i^l = \left(\sum_{j=1}^{s_l+1} W_{ji}^l\,\delta_j^{l+1}\right)f'\left(z_i^l\right) \quad (9)$$

- Determine the target partial derivatives as:

$$\frac{\partial J(W,B;x,y)}{\partial W_{ij}^l} = a_j^l\delta_i^{l+1} \quad (10)$$

$$\frac{\partial J(W,B;x,y)}{\partial B_i^l} = \delta_i^{l+1} \quad (11)$$

Via continuing these processes of the SGD, the objective factor $J(W, B)$ is reduced and so the DNN is trained. In this presented system, the DNN-based ensemble model is adopted. An entire framework is portrayed in Fig. 2.

In the initial level, the considered database D is split into 5 sub-collections: $D_1, D_2, \ldots, D_5$, where $D_i = \{x_i, y_i\}, i = 1, \ldots, 5$ includes the annotated instances acquired based on an equal allocation. In the initial cycle, the mixture of D_2, D_3, D_4 and D_5 is utilized as the learning collection and $D_1 = \{x_1, y_1\}$ is utilized as the test collection. For an input x_1, 4 classifiers in this level recommend respective $h_1(x_1), h_2(x_1), \ldots, h_5(x_1)$, wherein $h_i(x_i)$ denotes a binary factor and the subscript i of $h_i(x_1)$ is defined as i^{th} model.

Once the classifications in the initial round are completed, the predictions of all models are collected into $H_1 = \{h_1(x_1), h_2(x_1), \ldots, h_5(x_1)\}$ that is fused with the respective tag y_1 to create a fresh database D_1' for utilization in the 2nd level. Moreover, this process is continued for 5 cycles and 5 new databases are obtained such as $D_1', D_2', \ldots, D_5'$, where $D_i' = \{H_i, y_i\}, i = 1, \ldots, 5$. In the second level, the DNN is applied as the ensemble model. To predict the labels for unannotated instances, a 5-layer DNN is used. The input unit has 5 neurons that define attributes in the fresh database. In the hidden units, multiple nodes are considered for better prediction. The output unit has single neuron whose result is the label of corresponding instance.

Compared to the weighted mean and majority voting schemes in common ensemble mechanism that simply considers the linear correlations amid categorizers and demand for physical engagement,

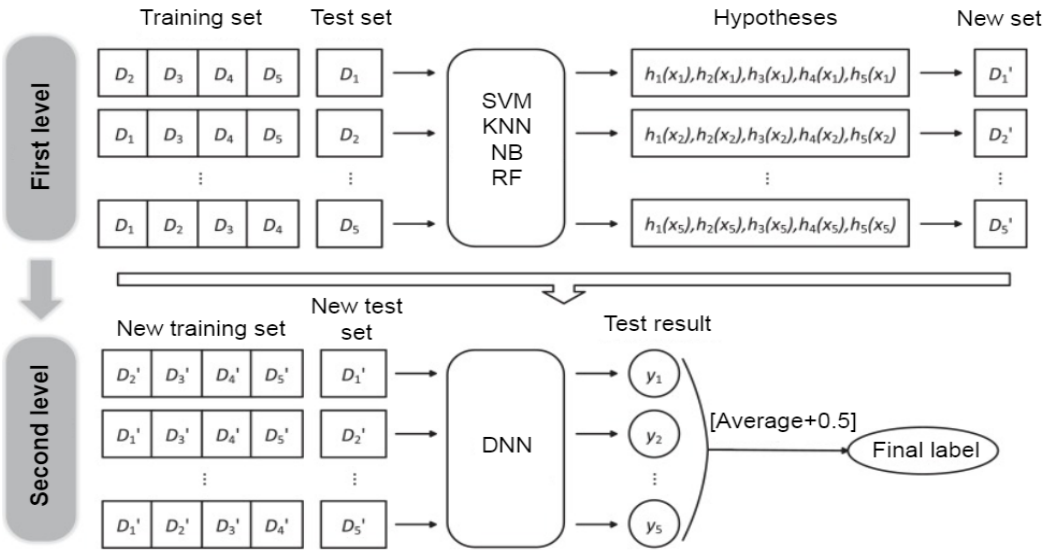

Figure. 2 Framework of DELM

the DELM trains the correlations automatically. The correlations between many categorizers and the tags of test instances are unrecognized and the consistency of the estimation is not ensured when merely a basic linear correlation is considered. But, the DELM utilized in the second level can automatically train intricate correlations, particularly nonlinear correlations. So, the DELM can create complete utilization of the details given by the information and ensure the estimation outcomes.

As depicted in Algorithm 1, this model is executed as the following:

1. *Step 1*: GAN generates synthetic data with an equal quantity by accepting an annotated collection as input. This type of synthetic data creation not merely fills the annotated collection, yet balances for the unbalanced labels of annotated collection. After that, DELM-based base learner is trained on annotated samples and synthetic samples to determine the tags of the unannotated collection. The unannotated sample having an equal pseudo tags determined using the presented support learner is then included to the annotated collection. This process is continued until the augmentation of annotated collection terminates. At this condition, the unannotated collection is rejected when it is not unfilled.
2. *Step 2*: Via considering the augmented annotated collection as input, GAN is utilized another time to create an equal quantity of synthetic data as augmented annotated collection. After that, the absolute categorizer is trained on both augmented annotated and synthetic collections.

3.2 CDSS based on DELM-GAN-based semi-supervised training algorithm

A considerable volume of patient data is acquired from the publicly available resources, providing a great base for data-driven CDSS performance. This CDSS collects clinical information from databases and facilitates physicians' treatments to deliver an accurate prognosis.

Algorithm 1 for semi-supervised training based on DELM-GAN

Input: $X = X_{Label} \cup X_{Unlabel}$, DELM-based base learner, GAN

Output: Final classifier C

Initialize: $X_L = X_{Label}, X_U = X_{Unlabel}, X_P = \emptyset$

$\boldsymbol{while}(|X_P|\ is\ modified)$

$X_G \leftarrow GAN(X_L)$;

Train DELM base learner on $X_L + X_G$;

Predict class labels L_{model} of X_U by applying DELM base learner;

$\boldsymbol{for}(each\ instance\ in\ X_U)$

Allocate the pseudo tag to unannotated example i;

Include $X_U(i)$ to X_P;

$\boldsymbol{end\ for}$

$X_L \leftarrow X_L + X_P$; //X_P: Pseudo-annotated databases

$X_U \leftarrow X_U - X_P$;

$\boldsymbol{end\ while}$

$X_G \leftarrow GAN(X_L), X_L \leftarrow X_L + X_G$;

Train final classifier C on X_L;

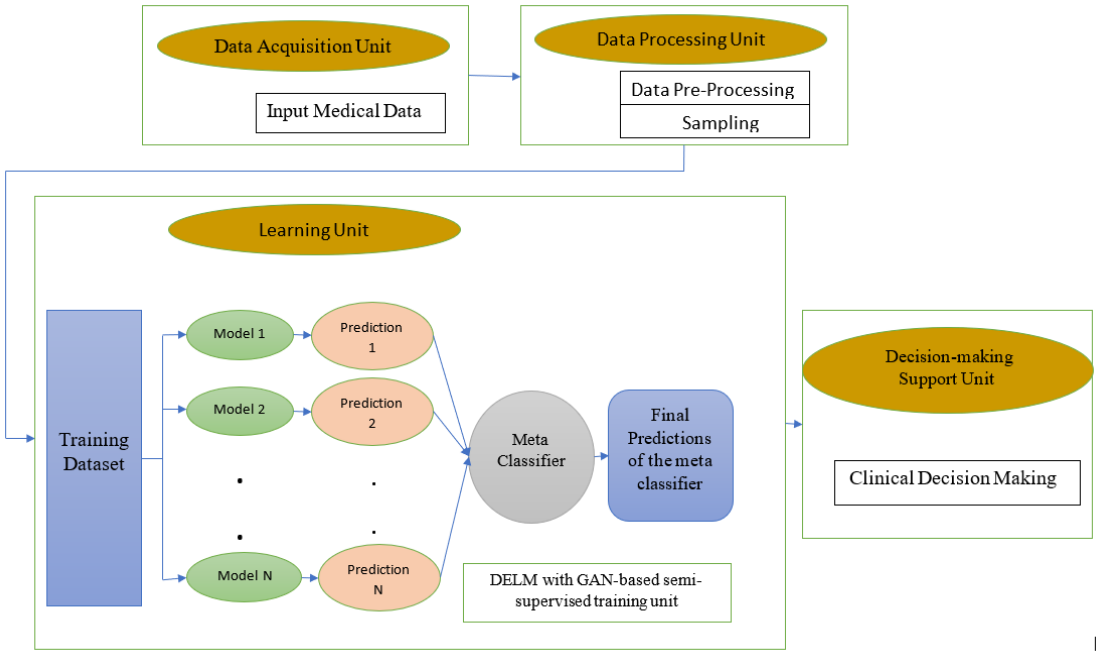

Figure. 3 Framework of CDSS based on presented DELM with semi-supervised GAN model

As depicted in Fig. 3, the DELM with GAN will be extended into the framework of CDSS, which is primarily developed to establish a strong execution on the clinical database with very tiny fraction of annotated collection and unbalanced labels. The whole model is split into 4 units:

1. Data collection unit: Initially, different clinical databases are collected from the freely available online resources.
2. Data processing unit: The collected data is given to the data processing unit, which aims to balance the target clinical database in the existence of unbalanced labels and a tiny part of annotated collection using DELM-GAN. After that, the attribute selection is performed for the pre-processed database.
3. DELM with GAN-based semi-supervised training unit: It learns the DELM with GAN-based semi-supervised training algorithm on the pre-processed clinical records having both annotated and unannotated collections.

Decision-making assistance unit: The well-learned semi-supervised trainer conducts a categorization process, where an unknown data like a patient is allocated via the tag, which estimates the unknown membership in one of the deliberate labels defining probable prognosis of multiple disorders.

4. Results and discussion

In this part, the efficiency of the presented DELM-GAN-based semi-supervised training algorithm is evaluated by implementing it in Python. In this experiment, 2 clinical record databases (e.g., SPECT and WDBC) and 1 clinical text database (Hallmarks) are considered. The SPECT database [26] contains 267 instances (110 positives and 157 negatives) with 22 attributes, whereas the WPBC database [27] contains 198 instances (47 positives and 151 negatives) with 34 attributes. Also, the clinical text database such as Hallmarks has a sentence length of 833, database size of 8474 and vocabulary size of 29141 for 3 different classes. Also, a comparative analysis is presented for proposed and existing algorithms: GAN-based semi-supervised training algorithm [15], Fuzzy_AHP+ANN [17], PSO-DNN [18], HTM+LSTM [21], DBSCAN+SMOTE-ENN+XGBoost [22] and LightGBM [25] regarding the following metrics:

- Accuracy: It is the proportion of exact classification over the total instances analyzed.

$$Accuracy = \frac{True\ Positive\ (TP)+True\ Negative\ (TN)}{TP+TN+False\ Positive\ (FP)+False\ Negative\ (FN)} \quad (12)$$

In Eq. (12), the quantity of positive tags exactly

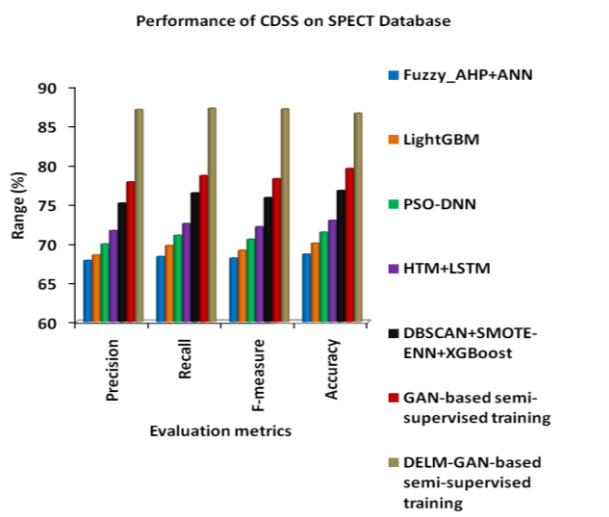

Figure. 4 Comparison of CDSS using different algorithms on SPECT database

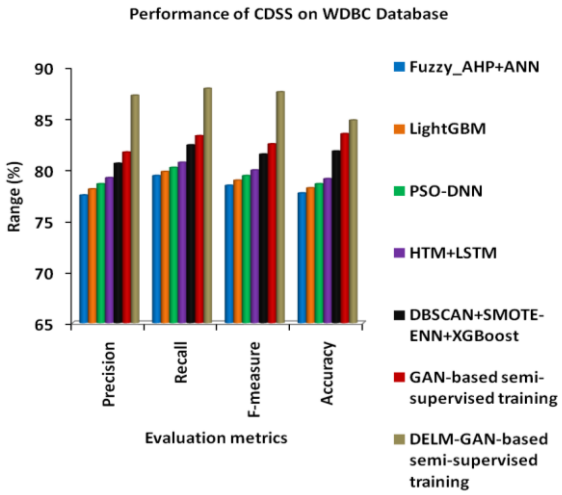

Figure. 5 Comparison of CDSS using different algorithms on WDBC database

categorized as positive (no disease) is TP, while the quantity of negative tags exactly categorized as negative (disease) is TN. In addition, FP is the quantity of negative tags categorized as positive, whereas FN is the quantity of positive tags categorized as negative.

- Precision: It determines the correctly classified class labels at TP and FP rates.

$$Precision = \frac{TP}{TP+FP} \tag{13}$$

- Recall: It is the proportion of class labels, which are correctly classified at TP and FN rates.

$$Recall = \frac{TP}{TP+FN} \tag{14}$$

- F-measure: It is calculated by

$$F-measure = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{15}$$

Fig. 4 displays the efficiencies of different prediction algorithms on the SPECT database. From this analysis, it is indicated that the DELM-GAN-based semi-supervised training algorithm achieves a higher efficiency on label prediction and proper diagnosis of multiple diseases compared to the other algorithms. The accuracy of the DELM-GAN-based semi-supervised training algorithm on the SPECT database is 26.15 % greater than the Fuzzy_AHP+ANN, 23.62 % greater than the LightGBM, 21.2 % greater than the PSO-DNN, 18.71% greater than the HTM+LSTM, 12.83 % greater than the DBSCAN+SMOTE-ENN+XGBoost and 8.86 % greater than the GAN-based semi-supervised training algorithms. Similarly, the precision, recall and f-measure of the DELM-GAN-based semi-supervised training algorithm are 21.21 %, 19.84 % and 20.46 % higher than all other classical algorithms on the SPECT database for CDSS. This is because of ensembling multiple distinct classifiers than every classifier functioning alone, as well as, it ability to learn and obtain hidden structures automatically.

Fig. 5 depicts the efficiencies of different prediction algorithms on the WDBC database. This analysis observes that the DELM-GAN-based semi-supervised training algorithm accomplishes a better prediction performance than all other algorithms. The accuracy of the DELM-GAN-based semi-supervised training algorithm on the WDBC database is 9.18 % higher than the Fuzzy_AHP+ANN, 8.48 % higher than the LightGBM, 7.93 % higher than the PSO-DNN, 7.24 % higher than the HTM+LSTM, 3.7 % higher than the DBSCAN+SMOTE-ENN+XGBoost and 1.59 % higher than the GAN-based semi-supervised training algorithms. Also, the precision, recall and f-measure of the DELM-GAN-based semi-supervised training algorithm are 10.1 %, 8.6 % and 9.32 % superior to the other classical algorithms on the WDBC database for CDSS. This is owing to the ability of DELM to predict the unknown correlations among classifiers and the label of samples than the single classifier models.

Fig. 6 illustrates the effectiveness of different prediction algorithms on the Hallmarks database. This analysis defines that the DELM-GAN-based semi-supervised training algorithm establishes

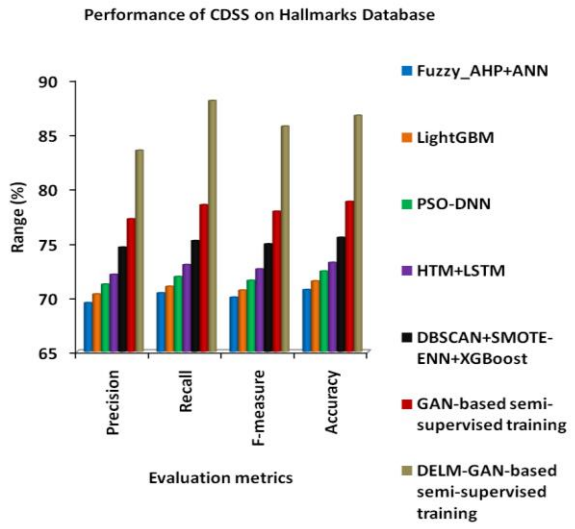

Figure. 6 Comparison of CDSS using different algorithms on Hallmarks database

higher prediction efficiency compared to all other algorithms. The accuracy of the DELM-GAN-based semi-supervised training algorithm on the Hallmarks database is 22.66 % larger than the Fuzzy_AHP+ANN, 21.29 % larger than the LightGBM, 19.78 % larger than the PSO-DNN, 18.47 % larger than the HTM+LSTM, 14.86 % larger than the DBSCAN+SMOTE-ENN+XGBoost and 10.05 % larger than the GAN-based semi-supervised training algorithms. Also, the precision, recall and f-measure of the DELM-GAN-based semi-supervised training algorithm are 15.21 %, 20.11 % and 17.55 % larger than the other classical algorithms on the Hallmarks database for CDSS. This is due to the ensemble of various classifiers, which takes the benefits of each classifier completely to achieve better prediction efficiency.

5. Conclusion

In this manuscript, the DELM with a GAN-based semi-supervised training algorithm was presented for CDSS. First, the publicly available clinical databases were acquired and pre-processed. Then, 2 levels of predictions were performed such as the initial level and the second level. In the initial level of prediction, SVM, KNN, NB and RF classifiers were trained on both annotated and unannotated databases separately to predict the label of the unannotated instances. Such independent predictions were ensemble by the DNN to get the absolute prediction and create the final classifier for robust CDSS. At last, the experimental results exhibit that the DELM with a GAN-based semi-supervised training algorithm has greater efficiency in predicting the label and supporting clinicians to provide a proper diagnosis for multiple diseases compared to the other classical algorithms. The DELM with a GAN-based semi-supervised training algorithm on SPECT, WDBC and Hallmarks database attains accuracies of 86.54 %, 84.83 % and 86.72 %, respectively, compared to the other existing prediction algorithms.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization, Thanabal; Methodology, Periya Nayaki Software, Simulation, Periya Nayaki; Writing-Original draft preparation, Periya Nayaki; Visualization, Investigation, Supervision, Thanabal; Reviewing and Editing, Thanabal.

References

[1] J. Li, J. Dai, A. Issakhov, S. F. Almojil, and A. Souri, “Towards Decision Support Systems for Energy Management in the Smart Industry and Internet of Things”, *Computers & Industrial Engineering*, Vol. 161, pp. 1-10, 2021.

[2] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, “An Overview of Clinical Decision Support Systems: Benefits, Risks, and Strategies for Success”, *NPJ Digital Medicine*, Vol. 3, No. 1, pp. 1-10, 2020.

[3] M. R. Davahli, W. Karwowski, and R. Taiar, “A System Dynamics Simulation Applied to Healthcare: A Systematic Review”, *International Journal of Environmental Research and Public Health*, Vol. 17, No. 16, pp. 1-27, 2020.

[4] A. Bashiri, B. A. Savareh, and M. Ghazisaeedi, “Promotion of Prehospital Emergency Care Through Clinical Decision Support Systems: Opportunities and Challenges”, *Clinical and Experimental Emergency Medicine*, Vol. 6, No. 4, pp. 288-296, 2019.

[5] T. Harada, T. Miyagami, K. Kunitomo, and T. Shimizu, “Clinical Decision Support Systems for Diagnosis in Primary Care: A Scoping Review”, *International Journal of Environmental Research and Public Health*, Vol. 18, No. 16, pp. 1-14, 2021.

[6] A. Awaysheh, J. Wilcke, F. Elvinger, L. Rees, W. Fan, and K. L. Zimmerman, “Review of Medical Decision Support and Machine-Learning Methods”, *Veterinary Pathology*, Vol. 56, No. 4, pp. 512-525, 2019.

[7] P. Shah, F. Kendall, S. Khozin, R. Goosen, J. Hu, J. Laramie, ... and N. Schork, "Artificial Intelligence and Machine Learning in Clinical Development: A Translational Perspective", *NPJ Digital Medicine*, Vol. 2, No. 1, pp. 1-5, 2019.

[8] Q. D. Buchlak, N. Esmaili, J. C. Leveque, F. Farrokhi, C. Bennett, M. Piccardi, and R. K. Sethi, "Machine Learning Applications to Clinical Decision Support in Neurosurgery: An Artificial Intelligence Augmented Systematic Review", *Neurosurgical Review*, Vol. 43, No. 5, pp. 1235-1253, 2020.

[9] N. P. Smadja, T. M. Rawson, R. Ahmad, A. Buchard, P. Georgiou, F. X. Lescure, ... and A. H. Holmes, "Machine Learning for Clinical Decision Support in Infectious Diseases: A Narrative Review of Current Applications", *Clinical Microbiology and Infection*, Vol. 26, No. 5, pp. 584-595, 2020.

[10] M. Fernandes, S. M. Vieira, F. Leite, C. Palos, S. Finkelstein, and J. M. Sousa, "Clinical Decision Support Systems for Triage in the Emergency Department using Intelligent Systems: A Review", *Artificial Intelligence in Medicine*, Vol. 102, pp. 1-56, 2020.

[11] J. E. V. Engelen and H. H. Hoos, "A Survey on Semi-Supervised Learning", *Machine Learning*, Vol. 109, No. 2, pp. 373-440, 2020.

[12] P. Kaur and A. Gosain, "Comparing the Behavior of Oversampling and Undersampling Approach of Class Imbalance Learning by Combining Class Imbalance Problem with Noise", In: *ICT Based Innovations, Springer*, Singapore, pp. 23-30, 2018.

[13] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results," In: *Proc. of 11th IEEE International Conf. on Information and Communication Systems*, pp. 243-248, 2020.

[14] A. Fernández, S. Garcia, F. Herrera, and N. V. Chawla, "SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary", *Journal of Artificial Intelligence Research*, Vol. 61, pp. 863-905, 2018.

[15] Y. Yang, F. Nan, P. Yang, Q. Meng, Y. Xie, D. Zhang, and K. Muhammad, "GAN-based Semi-Supervised Learning Approach for Clinical Decision Support in Health-IoT Platform", *IEEE Access*, Vol. 7, pp. 8048-8057, 2019.

[16] E. Alickovic and A. Subasi, "Medical Decision Support System for Diagnosis of Heart Arrhythmia using DWT and Random Forests Classifier", *Journal of Medical Systems*, Vol. 40, No. 4, pp. 1-12, 2016.

[17] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, "An Integrated Decision Support System based on ANN and Fuzzy_AHP for Heart Failure Risk Prediction", *Expert Systems with Applications*, Vol. 68, pp. 163-172, 2017.

[18] S. K. Lakshmanaprabu, S. N. Mohanty, S. Krishnamoorthy, J. Uthayakumar, and K. Shankar, "Online Clinical Decision Support System using Optimal Deep Neural Networks", *Applied Soft Computing*, Vol. 81, pp. 1-10, 2019.

[19] A. Masood, P. Yang, B. Sheng, H. Li, P. Li, J. Qin, ... and D. D. Feng, "Cloud-based Automated Clinical Decision Support System for Detection and Diagnosis of Lung Cancer in Chest CT", *IEEE Journal of Translational Engineering in Health and Medicine*, Vol. 8, pp. 1-13, 2019.

[20] M. A. Basset, G. Manogaran, A. Gamal, and V. Chang, "A Novel Intelligent Medical Decision Support Model based on Soft Computing and IoT", *IEEE Internet of Things Journal*, Vol. 7, No. 5, pp. 4160-4170, 2019.

[21] N. O. E. Ganainy, I. Balasingham, P. S. Halvorsen, and L. A. Rosseland, "A New Real Time Clinical Decision Support System using Machine Learning for Critical Care Units", *IEEE Access*, Vol. 8, pp. 185676-185687, 2020.

[22] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, "HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System", *IEEE Access*, Vol. 8, pp. 133034-133050, 2020.

[23] F. Orciuoli, F. J. Orciuoli, and A. Peduto, "A Mobile Clinical DSS based on Augmented Reality and Deep Learning for the Home Cares of Patients Afflicted by Bedsores", *Procedia Computer Science*, Vol. 175, pp. 181-188, 2020.

[24] U. Taşkıran and M. Çunkaş, "A Deep Learning Based Decision Support System for Diagnosis of Temporomandibular Joint Disorder", *Applied Acoustics*, Vol. 182, pp. 1-12, 2021.

[25] J. D. Schwalm, S. Di, T. Sheth, M. K. Natarajan, E. O'Brien, T. McCready, and J. Petch, "A Machine Learning–Based Clinical Decision Support Algorithm for Reducing Unnecessary Coronary Angiograms", *Cardiovascular Digital Health Journal*, Vol. 3, No. 1, pp. 21-30, 2022.

[26] https://archive.ics.uci.edu/ml/datasets/spect+heart

[27] https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(Prognostic)