Iraqi Journal of Science, 2025, Vol. 66, No. 12, pp: 5777- 5788
DOI: 10.24996/ijs.2025.66.12.40

ISSN: 0067-2904

Evolutionary-Based Feature Selection and Deep Recurrent Neural Network for Healthcare in IoTs System

Maysaa I Abdulhussain Almulla Khalaf[1]*, Nasreen J. Kadhim[1], Tuhsin H. A. Al-Jebur[2]
[1]*Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq*
[2]*University of Middle Technical, Baghdad, Iraq*

Received: 1/7/2024 Accepted: 11/12/2024 Published: 30/12/2025

Abstract

This paper proposes an optimized Deep Recurrent Neural Networks (DRNNs) for health care in IoT systems by selecting effective features and effective weight initialization based on genetic algorithm, to overcome the obstacles to training deep recurrent neural networks with restricted training data in feature space with high-dimensionality, such as over-fitting and vanishing/exploding gradients. Genetic algorithm is adopted for feature selection then the initial weights of the feature extraction layers are utilized to train the DRNNs, with the weights from the first stage of the feature selection layer being fixed. To enhance the network structures and determine the learning parameters, 10-fold cross-validation is employed. Performance evaluation of the trained DRNNs is conducted using test datasets not seen during the cross-validation process. Experimental results have revealed the effectiveness and advantages of the suggested approach.

Keywords: genetic algorithm, effective weight initialization, feature selection, and deep learning

اختيار الميزات المعتمد على الخوارزميات التطورية والشبكة العصبية العميقة المتكررة للرعاية الصحية في أنظمة إنترنت الأشياء

ميساء ابراهيم عبدالحسين الملاخلف [1], نسرين جواد كاظم [1], تحسين حسين عبد اللطيف[2]
[1] قسم علوم الحاسوب، كلية العلوم، جامعة بغداد، بغداد، العراق
[2] الجامعة التقنية الوسطى، بغداد، العراق

الخلاصة

تُقترح هذه الورقة استخدام شبكات عصبونية عميقة متكررة (DRNNs) محسّنة للرعاية الصحية في نظام إنترنت الأشياء (IoTs) من خلال اختيار الميزات الفعالة وتهيئة الأوزان بشكل فعّال باستخدام خوارزمية جينية، بهدف التغلب على العقبات في تدريب الشبكات العصبونية العميقة المتكررة مع بيانات تدريب محدودة في فضاء الميزات ذات الأبعاد العالية، مثل الإفراط في التكيّف (over–fitting) واختفاء/انفجار التدرجات. تم اعتماد الخوارزمية الجينية لاختيار الميزات، ثم يتم استخدام الأوزان الأولية لطبقات استخراج الميزات لتدريب DRNNs، مع تثبيت الأوزان من المرحلة الأولى لطبقة اختيار الميزات. لتعزيز هيكلية الشبكة وتحديد معايير

*Email: Maysaa.i@sc.uobaghdad.edu.iq

التعلم، يتم استخدام التحقق المتقاطع بتقسيم البيانات إلى 10 أجزاء (10 .fold cross-validation)-يتم تقييم أداء DRNNs المدربة باستخدام مجموعات اختبار لم تُستخدم أثناء عملية التحقق المتقاطع. وقد أظهرت النتائج التجريبية فعالية ومزايا النهج المقترح.

1. Introduction

Within the computing empire, the Internet of Things (IoTs) is considered a speedily growing technology [1]. It refers to a network infrastructure where the communication of distinct computing devices can be performed and where the collection and exchange of data are enabled without demanding human intervention. Ease of access, energy efficiency, flexibility, and portability has made IoTs appropriate to several industries, such as healthcare and medical informatics [2]. In the healthcare area, IoTs play an essential role in the reduction of the load on more conventional healthcare infrastructure. Healthcare solutions founded on IoTs are used to track patient information in real-time. IoTs IoT systems have recorded vast popularity across the world for healthcare applications, for example, diabetes management. Its prominence has been extended from personal monitoring to expert clinical diagnosis.

A long time ago, researchers have been concentrating their work on the recognition and prediction of diverse diseases using Machine Learning (ML). Through ML, the systems can be capable of automated learning and improving from experience without explicit programming. ML methods are often categorized according to learning as supervised, unsupervised, semi-supervised, and reinforcement learning. In supervised learning, model relationships and dependencies between the input features and target expectation output are figured out. So, the relationships that were learned from the preceding datasets can be the basis for predicting the output values for new data. In unsupervised learning, the unlabeled data is trained with the model [1-3]. Semi-supervised learning falls between supervised and unsupervised learning; both unlabeled and labeled data are used . Rain forest learning algorithms interact with their environment through generating actions and then discovering errors or rewards [2 and 4].

In ML, a discipline called deep learning (DL) exists, constructed on the basis of the theory of artificial neural networks (ANN) [3-5]. Computers use these computational models (ANNs) to evaluate and learn from incoming data. Each layer in NN enables data to be transformed through a sequence of nonlinear operations to learn input data with complex representations. In recent times, DL has ascended to prominence as a greatly observable subfield in ML because of its extraordinary results achieved in various applications [6].
Several ML exemplars may make use of DL. Firstly, in supervised ML, a NN is trained for information prediction or classification through the analysis of labeled datasets. The inputs, in this case, are both the outcome variables and the input features [4 and 6]. Training a NN is performed for making predictions through learning to reduce the mistake or cost that the gap between the target value and the predicted one produces. Back propagation is a common name for this recurring process. Deep learning methods like CNNs and RNNs are used in many supervised tasks, comprising but not restricted to classification and recognition of pictures, language translation, and sentiment analysis [7 and 8]. Secondly, in unsupervised ML, training a NN is performed for recognizing patterns or clustering datasets with the absence of labeled data [9]. In this perspective, the nonexistence of target variables is realized. The computer must have the ability to discover by itself any concealed patterns in the dataset [10]. Deep learning types that find widespread usage in unsupervised tasks containing anomaly detection, dimensionality reduction, and grouping are auto-encoders and

generative models [7]. Thirdly, Reinforcement Machine Learning (RL) states a ML method where training an agent is performed to make optimal selections within a given environment for maximizing a reward signal [2 and 5]. Engaging the agent in environmental interaction is done through the process of executing actions and subsequently observing the rewards that arise [4]. Deep learning can attain policies that are defined as a group of actions aiming to optimize the overall reward collected over a definite period. Deep reinforcement learning techniques such as Deep Q networks and Deep Deterministic Policy Gradient (DDPG) are often used in the reinforcement of several tasks, comprising game playing and robotics [2-5].
The standard method for managing diabetes necessitates people actively taking blood glucose (BG) measurements several times during the day with a finger prick test-self monitoring of blood glucose [6]. The recent progression and adoption of incessant glucose monitoring (CGM) devices enable more frequent glucose measurements (every 5 minutes), proving to be an effective approach in managing blood glucose (BG) and consequently enhancing the outcomes of individuals in clinical trials [6-8].
Typically, unsupervised learning is essential for acquiring features, such as through methods like restricted Boltzmann machine (RBM) and sparse auto-encoder (SAE) [7]. Supervised learning is preferable in the context of classification tasks, employing classifiers like support vector machines or feed-forward neural networks. Effectively integrating supervised and unsupervised learning is a crucial challenge for deep learning success in the traditional classification of patterns.

This paper proposes an optimized deep recurrent neural network (Opt_RNN) for health care in IoT systems by selecting effective features and effective weight initialization based on Genetic Algorithm (GA), to overcome the obstacles to train deep recurrent neural networks with restricted training data in feature space with high-dimensionality, such as over-fitting and vanishing/exploding gradients. Experiments have been performed on three datasets, evaluating the proposed approach performance by comparing it with traditional machine learning approaches such as Rain Forest, Naïve Bayes, and Decision Tree.
The structure of this paper is in as follows: several works that are associated with the suggested work are reviewed in Section Two. Section Three describes the basic principles of genetic algorithm, deep recurrent neural networks, and the proposed method. The experimental results are presented and discussed in Section Four. Section five presents the conclusions and some future directions.

2. Review of Related Works

The recognition of human activity monitoring has become a significant point of interest among researchers. The researchers are actively engaged in the field of interconnected intelligent healthcare and its many applications. These applications encompass tasks such as identifying falls [8], detecting activities to enhance energy efficiency in homes or offices [9], monitoring chronic diseases, creating a system to identify motion disorders in Autism patients [11], and utilizing the Internet of Things (IoTs) technologies for various purposes [12–14].

Kumar and Vashist [13] explored the challenges and accomplishments in eye care, particularly within the Indian community. They emphasized the need for effective care and viable solutions for numerous eye diseases leading to potential vision loss in India by the year 2020. John et al. [15] introduced a model named Genetic Algorithm Wrapper-Based Feature Selection and Naive Bayes for Anomaly Detection Model (GANBADM) in a Fog Environment. This model reduces unnecessary attributes, thereby decreasing time

complexity. Simultaneously, it enhances the accuracy of prediction using the Security Laboratory Knowledge Discovery Dataset (NSL-KDD) [13].

A classification-based technique to forecast illnesses represents a supervised approach that integrates diverse mechanisms, including Random Forest, Decision Tree, Linear and Polynomial Regression, and other systems. This encompasses the application of a Support Vector Machine (SVM) according to Huang S. et al. [14], K-Nearest Neighbour as proposed by Parry R. et al. [16], and the Logistic Regression method. The Logistic Regression approach stands out for its improved efficiency in terms of accuracy and precision when compared to alternative classification models. Supervised approaches showcase optimal performance when dealing with recognized cases. The prediction of the accuracy of the outcome is directly tied to the size of the training set, necessitating significant computational efforts. Though, in specific instances, the method may deviate due to excessive training.

Deep learning (DL) [13 and 14] is a subset of Machine Learning widely applied for predicting type-2 diseases by analyzing blood glucose levels and spectrogram images derived from these levels. Additionally, DL models can be employed with tabular datasets like PIMA to forecast diabetes. Each layer in the DL model represents a level of acquired information, with the layer closest to the input capturing low-level data elements, while the layer nearest to the output demonstrates a higher level of discrimination with more concise concepts. Deep learning typically requires extensive data for accurate classification and substantial computational resources for processing [16]. However, a significant drawback of deep learning models lies in their non-interpretable decision mechanism, limiting the trustworthiness of the models.

Zhao and colleagues [17] explored various deep learning techniques applicable to both human and mobile activities, specifically focusing on wearable and sensor-based approaches that perform automatic feature extraction. These deep learning methods are broadly categorized into discriminative, generative, and hybrid models. Within the generative category, the study mentions Restricted Boltzmann Machines, sparse coding, deep mixture models, and auto-encoders. Discriminative methods include deep neural models, convolutional neural networks, recurrent neural networks, and hydrocarbons. However, there are ongoing research challenges in areas such as decision fusion based on deep learning embedded in mobile devices, transfer learning, and addressing issues related to class imbalance in data.

Pasluosta and collaborators [18] investigated the Internet of Health Things (IoHTs) concept, which involves utilizing interconnected devices to monitor and manage patients' physical conditions through consolidating and integrating data on vital signs within hospital settings. IoHTs comprise four key stages: data collection, storage, processing, and presentation. The advantages of IoHTs are dual fold, as it can decrease service idle times and achieve an efficient allocation of limited resources. However, the main drawbacks identified in this study are associated with the sheer complexity and volume of the gathered data, posing a challenging task for caregivers [18].
Riazul Islam and colleagues [19] examined different aspects of IoTs-based technologies within the healthcare domain, proposing diverse structures and platforms for medical networks that establish the foundation for IoTs and facilitate the transfer and collection of medical data. The advantages of medical services based on IoTs include cost reduction and improved quality of life. However, a drawback is that the healthcare tools of IoTs are

equipped with slow processors, and these devices are unable to carry out computationally expensive tasks.

This paper proposed the Optimized Recurrent Neural Networks (Opt_RNNs) model firstly, by employing the Genetic Algorithm for feature selection and weight initialization, and secondly, Deep Recurrent Neural Networks are trained with effective features selected from the first stage and weight initialization. This model is used in health care in IoTs to overcome the obstacles of training deep recurrent neural networks.

3. Evolutionary based feature selection

3.1 Feature selection: Preliminaries

Within the field of ML, feature selection is becoming a pertinent problem. Its efforts are to select sufficient, required, and small subset of features representative of the entire set, to reduce irrelevant and redundant information.

In recent times, an increase has happened in the concentration on artificial intelligence, in particular ML. This technology has attained a large success owing to the huge volumes of available data and the broad computing capacity currently existing. The high level of existing data has permitted ML algorithms to become more and more complicated, and they can be nourished from dissimilar sources. This can be a problem despite its great benefit, as the data may be irrelevant and redundant, producing learning errors. Based on the available data, a set of features describing the problem considered is delineated in this perspective. The large datasets may generate broad features set, and the reduction of non-relevant information is of the greatest significance. The problem of feature selection aims to find a subset of features that is representative of the entire dataset to reduce redundant and non-relevant information within the dataset to improve the performance of a classification algorithm. This problem is considered difficult as the search space is determined as 2^{n}, where n resembles the number of features that construct the dataset [21 and 22].

For the feature selection problem, the methods for solving it can be categorized into a) Filter methods in which identification of the optimal set of features is performed by concentrating on the problem specificities in the dataset without considering the used classification algorithm. b) Wrapper methods where the subset of features is adjusted continuously based on the training phase of the machine learning model. Metaheuristics, exhaustive selection, backward elimination, and forward selection are among its most well-known types. c) Embedded methods where the difficulties that face wrapper and filter methods were overwhelmed [22].

Metaheuristics stand out in the wrapper methods. They are general purpose algorithms that can generate a solution to diverse optimization problems with a small number of adjustments. Their behavior is stochastic, and their optimization process is achieved through the balance of exploring the search space and exploiting the promising region [23]. Metaheuristics may provide high-quality results with these features in a reasonable time.

3.1.1 The proposed model: definitions and formulations

Addressing feature subset selection has been accomplished as an optimization problem with a *single objective*. Projecting the planned subset $\bar{F}$ is completed in the light of the delineated problem as in the proposed model definitions illustrated in what follows:

Definition 1 (*Subset* $\bar{F}$). Let $f_i \in F$ be a feature to be selected as a candidate feature in $\bar{F}$.The Prediction power\ Accuracy stated by (Eq.1) in what follows should be *maximized*. Accuracy is computed as the ratio of cases that are properly expected to the whole number of cases in F.

$$Accuracy = \frac{T_{Positive}+T_{Negative}}{T_{Positive}+T_{Negative}+F_{Positive}+F_{Negative}} \quad (1)$$

Where $T_{Positive}$ represents the number of positive instances that are properly classified, $T_{Negative}$ is the number of negative instances that are properly classified, $F_{Positive}$ is the number of negative instances that are incorrectly classified as positive and $F_{Negative}$ is the number of positive instances that are incorrectly classified as negative.

Alternatively, *Candidate Subset length*, or quantitatively, the number of the candidate features selected by the proposed approach to be involved in the candidate subset $\bar{F}$ should be *reduced*. At this point, for formulating our proposal, modeling the *feature subset selection problem* will be performed through the following definition:

Definition 2 (*feature subset selection problem* FSS_{Opt}). Suppose that f_i is a feature in $\bar{F}$. Then, the absence or presence of f_i can be expressed by a binary decision variable $g_i \in \{0,1\}$ (see Eq.2). Now, let $G = \{g_i | 1 \leq i \leq n\}$ be a vector comprising n such decision variables associated with the n features. At this point, for a vector$G$, the *feature subset selection* problem is a maximization problem that considers maximizing $Accuracy$ and minimizing the number of the selected features contained in the candidate subset $\bar{F}$ (see Eq.3).

$$g_i = \{1 \; if \; f_i \in \bar{F} \;\; and \;\; 0 \; otherwise \;, \quad (2)$$

$$Maximize \;\; FSS_{Opt}(G) = \sigma \times Accuracy_{\,candidates} + (1-\sigma) \times \left(1 - \frac{\sum_{i=1}^{n_{candidates}} f_i g_i}{n}\right) \quad (3)$$

Where $n_{candidates}$ represents the total number of features in the candidate subset $\bar{F}$,

The single optimization based model is designed as a weighted equation composed of two terms representing increasing accuracy and length reduction. It aims to include a small representative subset $\bar{F}$ of the original set of features F that are capable of representing F where it aims to increase the accuracy with a reduced number of features in $\bar{F}$.

3.1.2 Proposed Features Selection Model Based on GA

For a given dataset F, suppose that it totally consists of n features. F can then be represented by $F = \{f_i | 1 \leq i \leq n\}$, in which n denotes the number of dissimilar features that F contains. The objective is to spawn a subset $\bar{F} \subset F$ while dealing with the challenges in what follows:

✓ ***Increasing algorithm's prediction power \ Accuracy:*** the selected features subset $\bar{F}$ should contain the most critical features of the dataset F.

✓ ***Minimizing Candidate Subset length:*** the number of candidate features to be included in $\bar{F}$ should be minimized.

3.1.3 The applied genetic algorithm

Encoding of each genotype solution is performed using binary encoding and considered by a vector with a fixed-length, n, in which each gene value indicates the presence or absence of its associated feature. Then, for the proposed GA, the Cartesian product of presence/ absence of all n features can represent the complete search space S_{Space}:

$$S_{Space} = \prod_{i=1}^{n} g_n(\{0,1\}) = 2^n \quad (4)$$

Consider a population ρ of $K \ll S_{Space}$ genotype solutions, $P_{1 \leq k \leq K} \in \rho$. Then, $\forall k \in \{1, \dots, K\} \; and \; \forall j \in \{1, \dots, n\}: P_k = (P_{k1}, P_{k2}, \dots, P_{kn}) \; s.t. P_{kj} \in \{0,1\}$. The proposed GA can be described as a process articulated in an iterative function $\Psi: \rho \rightarrow \rho'$ with $\Psi(\rho_i) = \rho_{i+1}$, where ρ_i is the population at iteration i. The population begins with an initial population ρ_0 formed as illustrated in Eq.4 and carries on until a maximum number of iterations i_{max} is got. Formally speaking:

$\forall i \in \{1, \dots, n\}: \rho_0 = (\rho_{01}, \rho_{02}, \dots, \rho_{0n})$

$s.t.\ \rho_0$ is initiated through applying the formula:
$\rho_{0i}\ initialization \Leftarrow\ initialization_{random}$
Wherein for gene $j\ that\ corresponds\ to\ f_j \in \rho_{0i,}\ j \in \{0,1\}$
At every iteration i, three key operators will be involved in the evolution function Ψ: selection, crossover, and mutation, where each is controlled by its corresponding control parameter. Formally speaking:

$$\Psi = s_{\Theta_s} \circ x_{\Theta_x} \circ p_{\Theta_p} \quad (5)$$

Through applying the selection operator, s_{Θ_s}, in which the good quality chromosomes are copied to the subsequent generation to improve the average quality of the population while eliminating chromosomes with bad quality. The tournament selection has been adopted in the proposed work where only the fittest individual among several randomly picked individuals is selected for the following generation. The control parameter Θ_s decides the tournament size, which refers to the number of randomly selected individuals.

The Uniform Crossover has been adopted in the proposed algorithm. According to this kind of crossover, every gene of the child chromosome is created by randomly selecting the corresponding gene from one of its parents. An equal chance is given for both parents to contribute to creating the chromosomes that are formed from them. The crossover rate is decided through the control parameter Θ_x.

The Flip Mutation has been applied to the proposed model. The mutation rate is decided through the control parameter Θ_p.

The best solution, P^*, can be chosen as the result of the maximization problem FSS_{Opt}. Formally stated as:

$$P^*: \Leftrightarrow \nexists P \in \rho_{i_{max}} | \ FSS_{Opt}(G_P) > FSS_{Opt}(G_{P^*}) \quad (6)$$

3.2 Deep Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are supervised machine learning models composed of artificial neural networks designed to recognize patterns in sequence data, such as natural language and time series data [24], as illustrated in Figure 1. Training an RNN involves using a training dataset to minimize the discrepancy between the predicted outputs and the actual target values (i.e., the loss value) by optimizing the network's weights [25 and 26]. Recurrent Neural Networks (RNNs) are a class of artificial neural networks designed to recognize patterns in data sequences, such as time series or natural language. Unlike traditional feed-forward neural networks, RNNs have connections that form directed cycles, allowing information to persist. This feature makes them well-suited for tasks where context and order are important [26-28].

3.2.1 Recurrent Neural Networks Architecture

Simple RNNs have three layers, which are input, recurrent hidden, and output layers, as presented in Figure 1a [27].

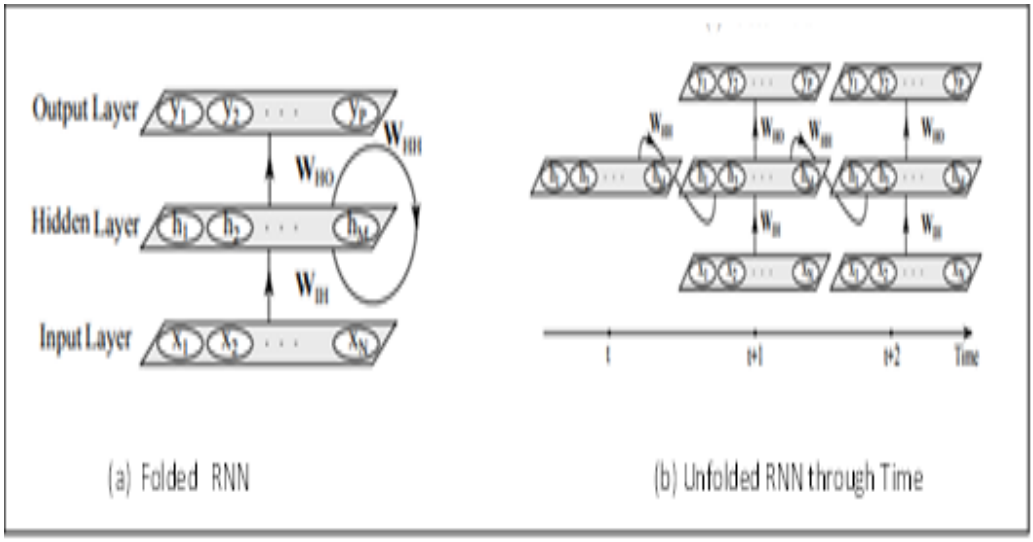

Figure 1: A simple Recurrent Neural Networks [27]

The input layer consists of N input units, which receive a sequence of vectors over time t, such as {..., xt−1, xt, xt+1, ...}, where each vector xt = (x1, x2, ..., xN). In a fully connected RNN, the input units are linked to the hidden units in the hidden layer via a weight matrix WIH. The hidden layer contains M hidden units, ht = (h1, h2, ..., hM), interconnected through time with recurrent connections, as shown in Figure 1b. Initializing the hidden units with small non-zero values can enhance the network's overall performance and stability [27 and 29]. The hidden layer represents the state space or "memory" of the system.

$$ht = fH(ot), \tag{7}$$

were

$$ot = WI\,H\,xt + WH\,H\,ht - 1 + bh \tag{8}$$

fH (·) is the hidden layer activation function, and bh is the bias vector of the hidden units. The hidden units are connected to the output layer with weighted connections W_{HO}.
The output layer has P units yt = (y1, y2, ..., yP) that are computed as

$$y\,t = fO\,(WH\,O\,ht + bo) \tag{9}$$

In the output layer, fO represents the activation functions, and bo is the bias vector. Given that input-target pairs are sequential over time, these steps are repeated over the time period t= (1,..., T). As indicated in Equations (7) and (9), an RNN comprises certain non-linear state equations that evolve over time. At each time step, the hidden states generate a prediction at the output layer based on the input vector. The hidden layer of an RNN consists of values that, independent of external factors, encapsulate all the necessary information about past states of the network over many time steps. This aggregated information determines the network's future behavior, enabling accurate predictions at the output layer [30]. The RNNs utilize a simple nonlinear activation function in each unit. Despite its simplicity, this structure can model complex dynamics if well-trained over multiple time steps.

3.3 Proposed Approach

Training deep recurrent neural networks with vast amounts of training data, particularly in input space of high-dimensionality, would be an issue of over-fitting. Though, in numerous applications, the necessitated massive amount of training data or deficient computer power for handling this data, as well as local minimum and vanishing/exploding gradient problems with inappropriate weight initialization.

Optimized Recurrent Neural Networks (Opt_RNNs) which utilizes a genetic algorithm for effective feature selection and weight initialization with deep recurrent neural network algorithm is proposed in this paper to overcome these issues as the following steps.

1. At the first stage, the datasets were pre-processed by removing reputation.
2. In the second stage, the Genetic Algorithm is adopted with random initial values for effective feature selection and weight initialization.
3. In the third stage, DRNNs are trained with suitable weight initialization and effective feature selection obtained from the second stage. Figure 2 illustrates the process.

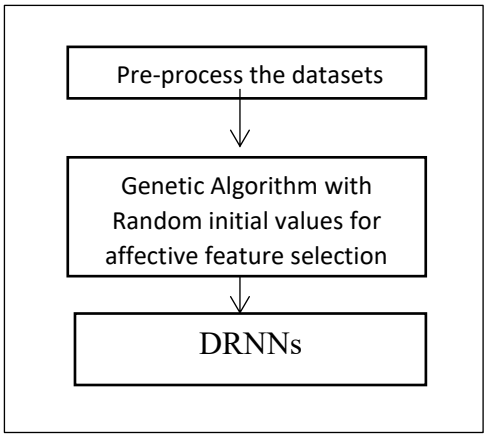

Figure 2: The Proposed Approach

4. Experimental Results and Discussion

4.1 Data Sets

In the experiments, three datasets were used. First, the diabetes-V1 dataset (https://archive.ics.uci.edu/dataset/34/diabetes) has 769 samples from two classes (500 /non-diabetes and 269 diabetes). Diabetes patient records were acquired from two sources: an automatic electronic recording device and paper records. Second, a diabetes_V2 dataset has 70694 samples from two classes (35346 diabetes-patients and 35348 non-diabetes-patients) from Kaggle Competition website (https://www.kaggle.com/wcukierski/datase). Third, a chronic disease dataset from the U.S. Department of Health & Human Services (https://resources.data.gov/schemas/dcat-us/v1.1/#accessLevel) that has 1048576 samples (patients).

4.2 Experiment Procedure

The dataset was divided into two sets: training and testing sets. Dividing the training set has been performed to produce a validation set and estimation set for cross-validation to determine the optimal or suitable network structure.

The proposed method (Opt_RNNs) was assessed with cross-validation, and each testing set was used only once for evaluating the proposed approach performance with the network structure trained using values selected by the cross-validation.

4.3 Results

For the three traditional methods (Decision tree, Naïve Bayes, and Random Forest), standard RNN, and the proposed Opt_RNN, the classification accuracies of cross-validation are illustrated in Tables 1 through 3. Figure 2 shows Comparison of three traditional methods, standard RNN and Opt_RNN in terms of testing accuracy.

Table 1: Performance evaluation on Diabetes V1

Methods	Train	Cross-Validation	Test
	Accuracy	Accuracy	Accuracy
Decision Tree	81.8%	82.5%	70.68%
Naïve Bayes	82.9%	83.9%	75.78%
Random Forest	83.2%	83.8%	76.40%
RNN	89.7%	89.9%	84.5%
Opt_RNN	95.3%	95.8%	94.9%

Table 2: Performance evaluation on Diabetes V2

Methods	Train	Cross-Validation	Test
	Accuracy	Accuracy	Accuracy
Decision Tree	79.9%	80.4%	70.58%
Naïve Bayes	81.5%	82.4%	74.68%
Random Forest	80.3%	80.7%	75.50%
RNNs	90.3%	90.9%	85.6%
Opt_RNNs	96.9%	95.7%	95.9%

Table 3: Performance evaluation on Chronic Disease.

Methods	Train	Cross-Validation	Test
	Accuracy	Accuracy	Accuracy
Decision Tree	78.8%	81.3%	82.58%
Naïve Bayes	82.4%	81.2%	78.8%
Random Forest	81.3%	80.2%	79.6%
RNNs	89.3%	89.9%	86.6%
Opt_RNNs	95.2%	94.3%	94.5%

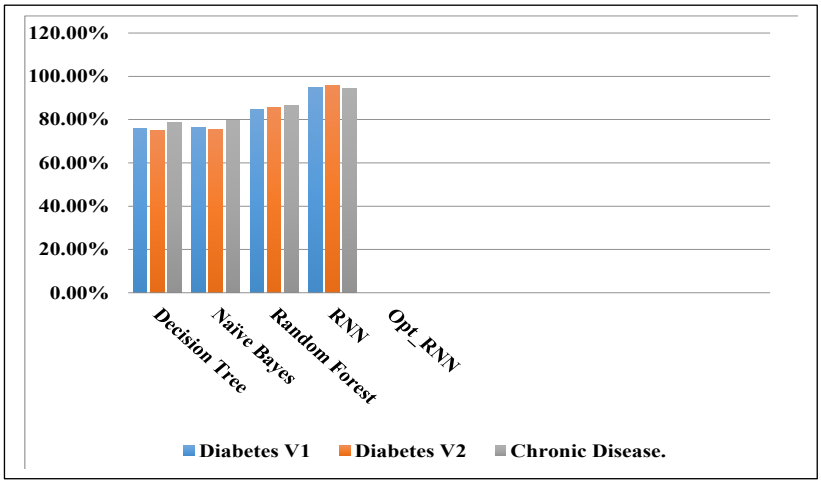

Figure 2: Comparison of three traditional methods, standard RNNs, and Opt_RNNs in terms of testing accuracy.

Statistical Significance Test: this paper applied a t-test as a statistical test on the three traditional methods: Decision Tree, Naïve Bayes, Random Forest, RNN, and our proposed method Opt_RNNs in terms of testing classification accuracy, to evaluate whether the performance differences among the methods are statistically significant. Table 4 shows the p-values from these tests, which explain that in terms of classification performance, Opt_RNNs significantly outperformed the three traditional methods and the RNN.

Table 1: Statistical test results (t-test).

Methods for comparison	p-value
Opt_RNNs vs. Decision Tree	2.1491e-07
Opt_RNNs vs. Naïve Bayes	3.4618e-07
Opt_RNNs vs. Random Forest	2.4738e-07
Opt_RNNs vs. RNN	0.0025

5. CONCLUSION

An Opt_RNNs approach is proposed in this paper to train deep recurrent neural networks with effective features and weight initialization based on GA. This approach can contest possible over-fitting and vanishing/exploding gradient problems in deep learning with restricted training data. The experimental results clarify that the proposed approach significantly outperforms the traditional approaches (Random Forest, Decision tree, and Naïve Bayes). Also, the preliminary experimental results of the standard recurrent neural networks have explained the benefits of the proposed method. More tests on this algorithm may be applied to deep neural networks for other applications as well might be directed in future investigations.

References

[1] A. Shah, et al. "Internet of things for sensing: A case study in the healthcare system," *Applied Sciences*, vol. 8.4, pp. 508, 2018.

[2] R. Arun Kumar, et al. "A survey of machine learning methods for IoT and their future applications." *Amity Journal of Computational Sciences* 2.2, pp 1-5, 2018.

[3] Z, Fotios, et al. "A review of machine learning and IoT in smart transportation." *Future Internet* 11.4, p 94, 2019.

[4] G. Umang, et al. "Analysis of machine learning algorithms for IoT botnet."*2nd International Conference for Emerging Technology (INCET)*. IEEE, 2021.

[5] A.U. Nabi, T. SNOUSSI, and Z.A. Abdalkareem. "*A Review of IoT Convergence in Healthcare and Smart Cities: Challenges, Innovations, and Future Perspectives*." Babylonian Journal of Internet of Things, pp.23-30, 2023.

[6] A. Rayan. "Identification and prediction of chronic diseases using machine learning approach." *Journal of Healthcare Engineering*, vol. 1, pp .2826127, 2022.

[7] A.I. Maysa, and GQ. John, "Class-specific pre-trained sparse autoencoders for learning effective features for document classification." *8th Computer Science and Electronic Engineering (CEEC)*. IEEE, 2016.

[8] L. Kezhi, et al. "Convolutional recurrent neural networks for glucose prediction." *IEEE Journal of Biomedical and Health Informatics* vol. 24, pp 603-613, 2019.

[9] Z. Wojciech, I. Sutskever, and O. Vinyals. "Recurrent neural network regularization." arXiv preprint arXiv, pp: 1409.2329, 2014.

[10] S.P. Naga, et al. "Using recurrent neural networks for predicting type-2 diabetes from genomic and tabular data." *Diagnostics,* vol. 12., pp. 3067, 2022.

[11] S.K, Rukiye, J.Q. Gan, and J.J Escobar. "A novel keyframe extraction method for video

classification using deep neural networks." *Neural Computing and Applications,* vol.35, p 24513-24524, 2023.
[12] Le., Yann, Y. Bengio, and G. Hinton. "Deep learning." *nature* 521.7553, p 436-444, 2015.
[13] V. Kumar. "Continuous glucose monitoring systems: a review." *Diagnostics* 3.4, p. 385-412, 2013.
[14] Yu. Huang and L. Zhao. "Review on landslide susceptibility mapping using support vector machines." *Catena,* vol. 165p. 520-529, 2018.
[15] O. John et al. "Genetic Algorithm based feature selection and Naïve Bayes for anomaly detection in fog computing environment*." Machine learning with applications, vol.* 6, pp.100156, 2021.
[16] M. Parry, et al. "k-Nearest neighbor models for microarray gene expression analysis and c-+ linical outcome prediction." *The pharmacogenomics journal,* vol. 10.4, p. 292-309 2010.
[17] Z. Huan, et al. "Wearable sensors and features for diagnosis of neurodegenerative diseases: A systematic review." *Digital Health,* vol. 9, 2023.
[18] C.F. Pasluosta, H. Gassner, J. Winkler, J. Klucken and B.M. Eskofier, "An Emerging Era in the Management of Parkinson's Disease: Wearable Technologies and the Internet of Things." *IEEE Journal of Biomedical and Health Informatics*, vol. 19, p. 1873- 1881, 2015.
[19] I. Riazul, et al. "The internet of things for health care: a comprehensive survey." *IEEE access*, vol. 3, P. 678-708, 2015.
[20] V. Kumar. "A review of microcantilevers for sensing applications." J. of Nanotechnology, vol. 3, p. 1-18,2007.
[21] I. Abubakr, Y.H. Ali, and T. A. Rashid. "Review on Hybrid Swarm Algorithms for Feature Selection." *Iraqi Journal of Science*, P. 5331-5344, 2023.
[22] A.I. Maysa, and G.Q. John. "Hybrid approaches to feature subset selection for data classification in high-dimensional feature space." *Artificial Intelligence Research*, vol. 9, p. 10, 2020.
[23] D.A. Mohammed, and J.K. Nasreen. "Extractive Multi-Document Summarization Model Based on Different Integrations of Double Similarity Measures." *Iraqi Journal of Science*, pp. 1498-1511, 2020.
[24] M. Almaiah, et al. "Detecting DDoS attacks using machine learning algorithms and feature selection methods." *International Journal of Data and Network Science* 8.4, pp. 2307-2318, 2024.
[25] A.I. Maysa, and G.Q. John. "Deep Classifier Structures with Autoencoder for Higher-level Feature Extraction." *IJCCI*, 2018.
[26] M.M. Elijah, and Z. Wang. "Hybrid wrapper feature selection method based on genetic algorithm and extreme learning machine for intrusion detection." *Journal of Big Data* 11.1pp. 24, 2024.
[27] S. Hojjat, et al. "Recent advances in recurrent neural networks." arXiv preprint arXiv:1801.01078, 2017.
[28] M.R. Larry R, and L. Jain. "Recurrent neural networks." *Design and Applications* 5.64-67, pp. 2, 2001.
[29] G. Stephen. "Recurrent neural networks." *Scholarpedia* 8.2, p. 1888, 2013.
[30] A.l. Mohammed Ali, et al. "A survey of machine and deep learning methods for internet of things (IoT) security." *IEEE communications surveys & tutorials* 22.3, p. 1646-1685, 2020.