INTELLIGENT DEEP LEARNING BASED PREDICTIVE MODEL FOR CORONARY HEART DISEASE AND CHRONIC KIDNEY DISEASE ON PEOPLE WITH DIABETES MELLITUS

A. Thasil Mohamed [1*], Sundar Santhoshkumar [2], Vijayakumar Varadarajan [3]

[1]Doctoral Research Scholar, Department of Computer Science, Alagappa University, Karaikudi, 630 003, India

[2]Assistant Professor, Department of Computer Science, Alagappa University, Karaikudi, 630 003, India

[3]School of Computing Science and Engineering, The University of New South Wales, Sydney, Australia

Email: ameerthasil@gmail.com[1*] (corresponding author), santhoshkumars@alagappauniversity.ac.in[2], vijayakumar.varadarajan@gmail.com[3]

DOI: https://doi.org/10.22452/mjcs.sp2022no1.7

ABSTRACT

Presently, process analytics extracts the knowledge from the past data to explore, monitor, and improve the processes. The recently developed deep learning (DL) models find it helpful to analyse medical data and make decisions. Among various diseases, type 2 diabetes mellitus (T2DM) becomes a widespread disease over the globe and it leads to severe outcomes. Chronic kidney disease (CKD) and coronary heart disease (CHD) are the major illness occurred in people with T2DM. Since the earlier prediction of the risk factors related to CKD and CHD on T2DM persons is necessary, this study focuses on the design of intelligent feature selection with deep learning based risk factor prediction (IFS-DLRFP) model. The proposed IFS-DLRFP technique intends to determine the early warning to the patients with T2DM to develop CKD or CHD. In addition, the IFS-DLRFP technique includes the design of fruit fly optimization algorithm (FFOA) based feature selection technique to choose an optimal set of features. Moreover, firefly optimization with gated recurrent unit (FF-GRU) based classification technique is derived to allocate appropriate class labels to the input data. The FF-GRU technique performs the hyperparameter tuning process using FF technique. In order to ensure the better performance of the IFS-DLRFP technique, a wide range of simulations take place on benchmark datasets and the simulation outcomes reported the supremacy of the IFS-DLRFP approach over the recent techniques.

Keywords: ***Risk Factors, Diabetes Mellitus, Prediction Model, Deep Learning, Feature Selection, Parameter Tuning, Chronic Kidney Disease, Coronary Heart Disease***

1.0 INTRODUCTION

Diabetes mellitus (DM) is a chronic disorder which takes place when the body cloud not produces sufficient or make efficient usage of insulin, and is induced by an environmental factor as well as genetic predisposition. 366 million persons have DM in 2011; half of this (183 million people) aren't diagnosed. The count of persons with global DM is increasing and in 2030 this has been increased to 552 million. The DM has been well-determined risk factor to cardiovascular disease (CVD). Patients with type 2 diabetes mellitus (T2DM) have high mortality and cardiovascular morbidity and are most severely affected by CVD than non-diabetic subjects [1]. Diabetic vascular disease has been accountable to two-four-fold increase in the existence of coronary artery disease (CAD) and stroke, and two-eight-fold increase under the risk of cardiac arrest. Consider the growing amount of global epidemic of T2DM and the cardiovascular event survivors, it is predicted the number of persons with T2DM at a high cardiovascular risk to increase, posing the biggest challenges for global healthcare system. Therefore, Cost-effective policies to reduce cardiovascular risk in this population are highly required [2]. The current review focus on the effect of risk factor in the cardiovascular risk worldwide, and the role of detecting silent ischemia and subclinical atherosclerosis in the asymptomatic person with diabetes. Current evidence indicates that CHD risk in T2DM isn't universally equivalent to the risk of person with earlier cardiovascular disease, but is very heterogeneous. A Meta-analysis of thirteen epidemiological analyses, including 45,108 persons with and without diabetes recognized that in T2DM without prior CHD, the CHD risk was 43% lower than persons without diabetes with an earlier myocardial infarction [3].

Chronic kidney disease (CKD) is known as a significant public health problem. CKD affects 10%-15% of adults globally and increases the risk of many medical adverse results [4]. CVD has become a major complication of CKD. In fact, up to 50% of patients with CKD pass away as a result of CVD even before many of them reached final stage of renal disease needing renal replacement therapy. Risk prediction is essential for making decisions for primary prevention of CVD like aspirin/statin therapy [5]. But, there are many key challenges in forecasting CVD risk in patients with CKD. Initially, conventional risk factors have been demonstrated to perform inefficiently in this medical population. Additionally, even though many researches have shown that patients with CKD are at great risk of CVD, remarkably, medical guidance is inconsistent about how to use data on CKD measures for forecasting CVD risk [6]. This review summarizes the present study about CVD predictions with CKD measure and CVD predictions including CKD, with emphasis on current research.

With the accessibility of bio-medical data, the usage of ML approaches in health care for improving disease predictive methods are developed more widespread [7]. Furthermore, techniques like DL and methods such as ensemble learning has considerably enhanced the prediction power of ML techniques. Through deriving features in Electronic Health Records (EHR), precise disease predictive methods could be established [8]. At the personal level, a doctor might determine the onset of CKD with laboratory test by looking at key parameters like albumin-creatinine ratio and the glomerular filtration rate (eGFR) [9]. At the same time, from the public healthcare point of view, laboratory information is generally inaccessible on a largescale. But, 2 kinds of information could typically be extracted from the insurance company's database: medications and diagnoses for every person visit at the hospital.

This study presents an intelligent feature selection with deep learning based risk factor prediction (IFS-DLRFP) model to determine the early warning to the patients with T2DM to develop CKD or CHD. Besides, the IFS-DLRFP technique includes the design of fruit fly optimization algorithm (FFOA) based feature selection approach for choosing an optimum set of features. Furthermore, firefly optimization with gated recurrent unit (FF-GRU) based classification technique is derived for assigning proper class labels to the input data. The FF-GRU technique carries out the hyperparameter tuning process by the use of FF technique. For assessing the effectual results of the IFS-DLRFP technique, a comprehensive experimental analysis is carried out on benchmark dataset and the simulation results stated the improved performance of the IFS-DLRFP technique compared to other techniques.

The rest of the paper is organized as follows. Section 2 offers the literature review and section 3 discusses the proposed model. Next, section 4 offers the performance validation and section 5 concludes the study.

2.0 RELATED WORKS

Hossain et al. [10] developed a risk predictive method utilizing administrative data which employs network based features and ML methods for measuring the risk of CVD in T2D persons. For that, 2 cohorts (viz., person with T2D and CVD and persons with only T2D) have been diagnosed from an administrative database gathered from the private health care funds depends on Australia. The 2 standard disease networks have been made from these 2 cohort studies. The last disease network was later made from 2 standard disease networks via normalization. Jamthikar et al. [11] proposed an office based CVD risk calculator with an ML model which used a focused carotid ultrasound. The proposal of this work has been separated into 2 stages. The initial stage involves gathering eighteen office based biomarkers comprising of 6 medical risk factors (sex, age, systolic blood pressure, body mass index, smoking, and diastolic blood pressure) and twelve carotid ultrasound image based phenotypes.

Kim and Kang [12] devised NN based predictions of CHD risk with feature correlation analyses (NN-FCA) utilizing 2 phases. Firstly, the feature selection phase that creates features allowing to the significance in forecasting CHD risk is ranked, and next, the feature correlation analyses phase, where one learned on the occurrence of correlations among feature relationships and the data of every NN predictor output has been established. Senan et al. [13] focus on calculating a dataset gathered from four hundred persons having twenty four features. The 4 classification methods used in this work are SVM, KNN, DT, and RF.

Song et al. [14] developed a new temporal enhanced GBM method which dynamically updates and ensemble learners according to a novel event in person timelines for improving the predictive performance of CKD amongst persons with diabetes. With a wide range of deidentified EHR information on a retrospective cohort of 14,039 adults with type 2 diabetes and GBM as the base learners, they have authenticated this presented Landmark Boosting method against 3 advanced temporal methods to roll forecasts of 1 year CKD risks.

Segal et al. [15] analyzed 10,000,000 medicinal insurance claims from 550,000 patients' records with commercial health insurance databases. They employed a feature embedding model according to the performance of Word2Vec model for additionally capturing temporal data for the 3 major elements of the data: medications, diagnosis, and

procedures. For the analyses, they employed XGBoost method. In Chimwayi et al. [16], neuro fuzzy technique is employed for determining the risk of CKD in person. Prediction made by the neuro fuzzy provided a 97% of accuracy. With the selected feature, predictions for CKD disease are made for identifying the risk. With hierarchical clustering, 3 clusters were made that show there are stronger relationships AMONG CHRONIC KIDNEY AND DIABETES.

3.0 THE PROPOSED MODEL

In this study, a novel IFS-DLRFP technique has been developed to determine the risk factors of CKD and CKD among T2DM patients. The proposed IFS-DLRFP technique encompasses four processes namely pre-processing, FFOA based FS, GRU based classification, and FF based hyperparameter optimization. Fig. 1 demonstrates the overall process of IFS-DLRFP model. The detailed working of these modules is given in the following sections.

3.1 Stage I: Data Pre-processing

Primarily, the input medicinal data has been pre-processed in 2 phases such as format conversion as well as missing value replacement. In the format conversion stage, the raw medicinal dataset has been converted as compatible. arff format. Afterward, the missing values to be in the dataset have been filled with utilize of median approach.

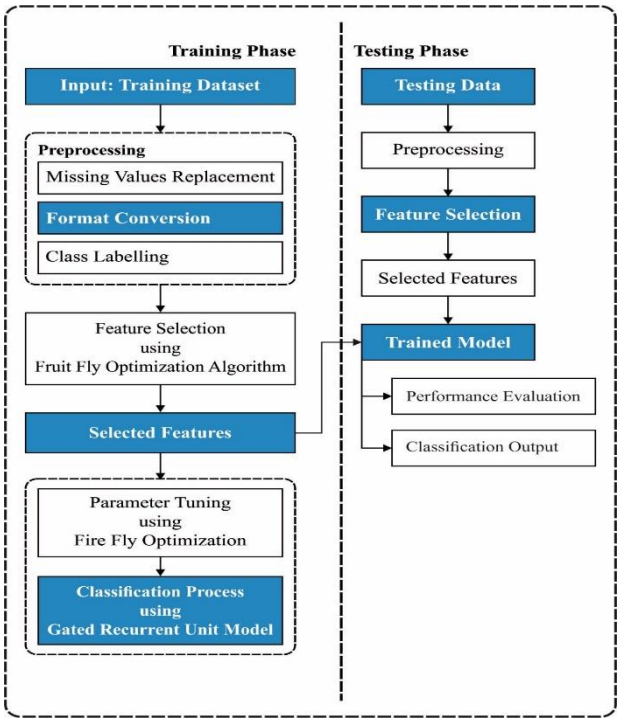

Fig. 1: Overall process of IFS-DLRFP model

3.2 Stage II: Process involved in FFOA-FS Technique

At the second stage, the pre-processed data is applied as input to the FFOA-FS technique to determine an optimum subset of features. FFOA is a novel heuristic approach which simulates the foraging activity of fruit fly in nature for seeking optimum solutions of the objective function. The foraging iteration model of fruit fly is the fundamental concept and the primary steps are given in the following:

Step 1: initiate parameter Set *Sizepop* and *Maxgen* of the population size, and initiate the population location:

$$(X_{-axis}, Y_{-axis}). \quad (1)$$

Step 2: fruit flies search in the olfactory scheme that could makes the search step and search direction arbitrarily. Random value (RV) is supposed that search distance, and the location of the population is upgraded at the same time:

$$\begin{cases} \mathrm{X_i} = \mathrm{X_{-axis}} + \mathrm{RV} \\ \mathrm{Y_i} = \mathrm{Y_{-axis}} + \mathrm{RV} \end{cases}. \quad (2)$$

Step 3: as the accurate position of the food isn't known, there is need for calculating the distance ($Dist_i$) among the origin of the coordinate and fruit flies [17], later evaluate the taste concentration variable (S_i):

$$Dist_i = \sqrt{\mathrm{X_i^2} + \mathrm{Y_i^2}} \quad (3)$$

$$S_i = \frac{1}{\mathrm{Dist_i}}. \quad (4)$$

Step 4: replace the fruit flies favour concentration definition value (S_i) to the fitness function, the taste concentration decision functions, then attain the separate taste concentration of the fruit fly $Smell_i$.

$$Smell_i = Fitness(S_i). \quad (5)$$

Step 5: recognize individuals with a maximum favour concentration in the drosophila population.

$$[bestSmell, bestIndex] = min\,(Smell). \quad (6)$$

Step 6: preserve the optimal favour coordinate and concentration value, as well as another individual in the population fly to this location:

$$SmellBest = bestSmell. \quad (7)$$

$$\begin{cases} \mathrm{X_{-axis}} = \mathrm{X\,(bestIndex)} \\ \mathrm{Y_{-axis}} = \mathrm{Y\,(bestIndex)} \end{cases} \quad (8)$$

Step 7: End criteria, determine either the concentration of the optimal location is superior to the prior generation, and attain the maximal amount of iterations; or else, skip step for to entering the iterative optimization.

FS has been assumed a binary optimized issue. It can utilize binary strings for representing the solution of FS issue. The vector has of d element, where d implies the amount of features under the original dataset. When it can choose the equivalent features before, fix them to "1" else, fix them to "0". The decision variable of issue are explained as follows:

$$X = (x_1, x_2, \dots, x_d), x_i \in [0,1], i = 1,2, \dots, d. \quad (9)$$

The FS has been assumed as multi-objective optimized issue. For maintaining the balance amongst the amount of FS and the classifier accuracy of solutions [18], the FF has been planned as follows:

$$Fitness = \alpha\gamma_R(D) + \beta\frac{|R|}{|C|} \quad (10)$$

γ_R signifies the classifier error rate of provided classifier (where GRU has been utilized). α demonstrates the weight of classification accuracy, and β refers the weight of feature decrease. $|R|$ stands for the amount of FS, $|C|$ demonstrates the entire number of features.

3.3 Stage III: Process involved in FF-GRU Technique

Finally, the FF-GRU technique is applied for the classification of input data. RNN is appropriate for nonlinear time series processing. The RNN includes output layer y, input layer x, and hidden layer h. While handling time series data, the RNN could be unfolded as right part. The output layer and hidden layer could be estimated based on Eqs. (11) and (12), correspondingly.

$$y_t = g(s_t * w_{hy}) \quad (11)$$

$$s_t = f(x_t * w_{sx} + s_{t-1} * w_{ss}) \quad (12)$$

In spite of their popularity as universal function approximate and easier execution, the RNN approach is confronted with gradient exploding or vanishing problems. During the trained phase of RNN, gradient is evaluated in the resultant to primary layers of RNN. When the gradient is less than 1, the gradient of the primary various layers would be small by using several multiplications. In contrast, the gradient would become larger when the gradient is greater than 1. Thus, sometimes it reasons the gradient for developing nearly zero or larger once it gains the first layer of RNN [19]. Therefore, the weight of the first layer won't obtain upgraded in the training phase. Thus, simple RNN could not be appropriate for complicated issues.

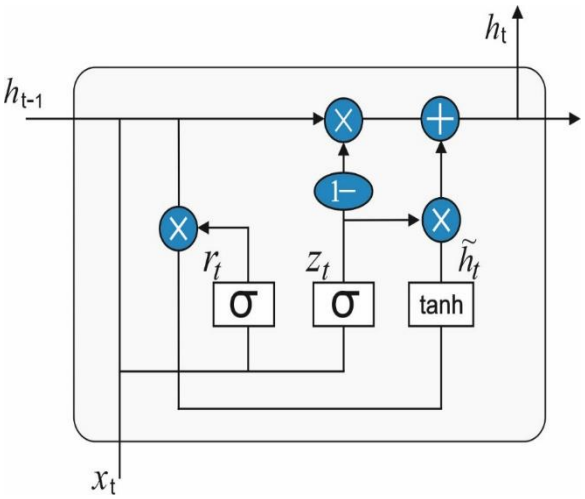

Fig. 2: GRU structure

During this work, a GRU based model has been presented for handling multivariate time series imagery data that is resolve the vanishing gradient problems of typical RNN. Fig. 2 illustrates the framework of GRU model. According to the prior output h_{t-1} and the present input x_t, a reset gate has been employed for determining which portion of data have to be reset, as estimated in Eq. (13), when an update gate is utilized for updating the output of GRU h_t, as estimated in Eq. (14). The candidate hidden layer has been evaluated based on Eq. (15). The present output could be attained based on Eq. (16). The gates, such as, z_t & r_t, and parameters, such as, W_z, W_r and W, of the GRU, has been upgraded under the trained phase.

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (13)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (14)$$

$$h_t^{'} = \tan h(W \cdot [r_t * h_{t-1}, x_t]) \quad (15)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * h_t^{'} \quad (16)$$

For optimally selecting the hyperparameters of the GRU model, the FF technique is applied and thereby improves the overall prediction performance.

The firefly algorithm (FA) has been presented as a metaheuristic approach that simulates the flashing performance of fireflies. The FA utilizes a collection of rules because the fireflies are unisex, hence some firefly attract another fireflies. As well, there is a relationship among brightness and attractiveness, in which the brighter fireflies would attract the less bright one. Where the attractiveness and brightness are improved by reducing the distance among the fireflies.

Generally, the attractiveness (β) among the *i*th firefly and *j*th firefly has been calculated as

$$\beta = \beta_0 \times e^{(-\gamma m^2)} \quad (17)$$

In which y represent the coefficient of light absorptions, and $\beta_0 = 1$ represent the attractiveness at the distance $m = 0$, whereas m represent the distance among the *i*th firefly X_i and the *j*th firefly X_j is calculated in the following equation:

$$m_{ij} = \|\mathrm{X_i} - X_j\| = \sqrt{\sum_{\mathrm{k}=1}^{d} \left(X_{ik} - X_{jk}\right)^2} \quad (18)$$

The effort of firefly i is involved to other brighter firefly j, is computed as [20]:

$$\mathrm{X_i} = X_i + \beta \times \left(X_i - X_j\right) + r_4 \times \varepsilon_i \quad (19)$$

Whereas $r_4 \in [0,1]$ represent an arbitrary number and $\varepsilon_i \in N(\mu, \sigma)$ represent an arbitrary vector. The fundamental step of the FA is shown in Algorithm 2.

Algorithm 1: Firefly algorithm

Create a collection of N fireflies $X_i (i = 1, 2, \ldots, N)$.
Calculate the fitness value for all the solutions X_i.
Light intensity I_i at x_i is defined as $f(x_i)$
Define the light absorption coefficients y
repeat
  for $i = 1$: N do
    for $j = 1$: i do
      if $f_i < f_j$ then
        Move the ith firefly X_i towards the jth firefly.
      end if
      Upgrade the attractiveness according to the distance m using exp $[-\gamma m]$.
      Upgrade the location of firefly by Eq. (19).
    end for
  end for
  Define the optimal solution.
until (k $< MaximumGenerations$)
Return the optimal solution.

4.0 EXPERIMENTAL VALIDATION

This section examines the performance of the IFS-DLRFP technique on the applied benchmark dataset [21], which comprises a total of 400 instances with 24 attributes and 2 classes. The results are investigated in various aspects. Fig. 3 showcases the confusion matrices produced by the IFS-DLRFP technique on the classification of DM. On the execution run-1, the IFS-DLRFP technique has categorized233 instances into absence of DM and 122 instances into presence of DM. Moreover, under run-3, the IFS-DLRFP technique has classified a total of 234 instances into absence of DM and 123 instances into presence of DM. Furthermore, under execution run-5, the IFS-DLRFP technique has categorized 235 instances into absence of DM and 125 instances into presence of DM.

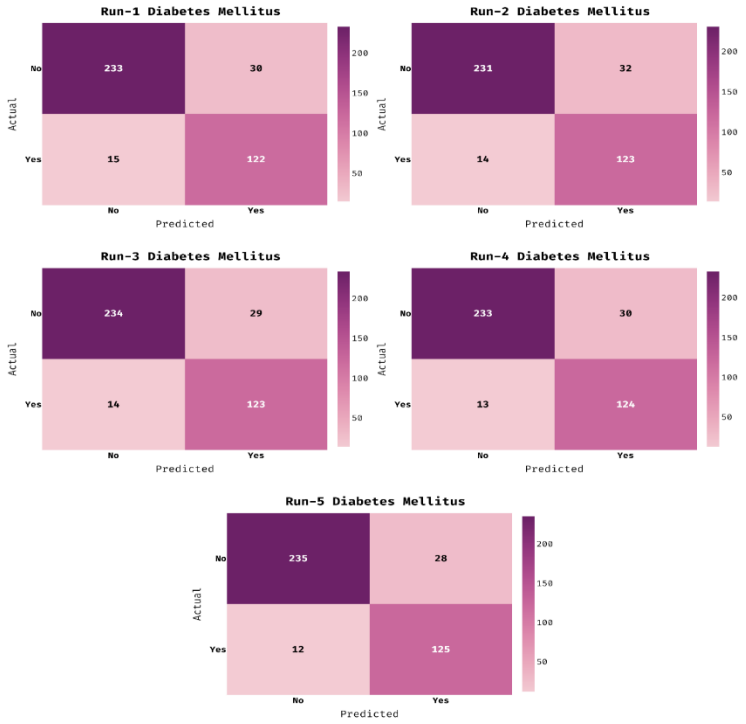

Fig. 3: Confusion matrix of IFS-DLRFP model on DM classification

Table 1 offers the DM classification outcomes analysis of the IFS-DLRFP technique under varying runs. The results depicted that the IFS-DLRFP technique has attained effective outcomes under every run. For instance, with run-1, the IFS-DLRFP technique has classified the DM with the pre_n, rec_l, $spec_y$, acc_y, and F_{score} of 0.9395, 0.8859, 0.8905, 0.8875, and 0.9119 respectively. Eventually, with run-3, the IFS-DLRFP method has classified the DM with the pre_n, rec_l, $spec_y$, acc_y, and F_{score} of 0.9435, 0.8897, 0.8978, 0.8925, and 0.9159 correspondingly. Meanwhile, with run-5, the IFS-DLRFP system has classified the DM with the pre_n, rec_l, $spec_y$, acc_y, and F_{score} of 0.9514, 0.8935, 0.9124, 0.9000, and 0.9216 respectively.

Table 1**:** Result analysis of IFS-DLRFP model on DM classification

No. of Runs	**Precision**	**Recall**	**Specificity**	**Accuracy**	**F-Score**
Run-1	0.9395	0.8859	0.8905	0.8875	0.9119
Run-2	0.9429	0.8783	0.8978	0.8850	0.9094
Run-3	0.9435	0.8897	0.8978	0.8925	0.9159
Run-4	0.9472	0.8859	0.9051	0.8925	0.9155
Run-5	0.9514	0.8935	0.9124	0.9000	0.9216
Average	**0.9449**	**0.8867**	**0.9007**	**0.8915**	**0.9149**

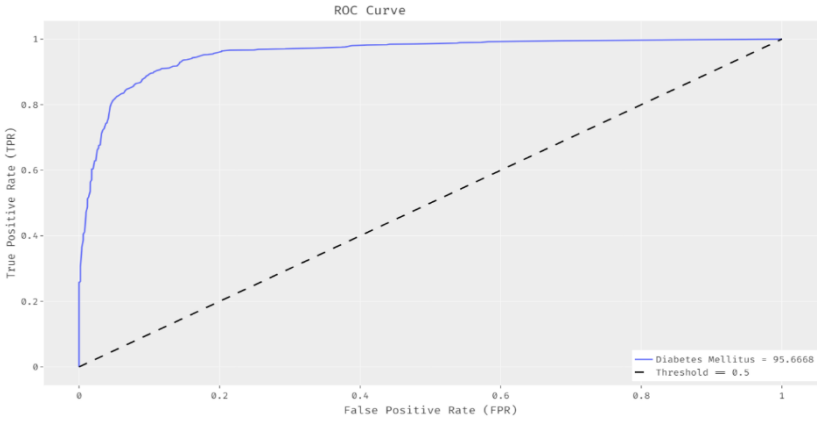

Fig. 4: ROC analysis of IFS-DLRFP model on DM classification

Fig. 4 defines the ROC analysis of the IFS-DLRFP technique on the classification of DM. The figure stated that the IFS-DLRFP technique has resulted in a higher ROC of 95.6668%.

Fig. 5 depicts the confusion matrices produced by the IFS-DLRFP system on the classification of CAD. On the execution run-1, the IFS-DLRFP approach has categorized 362 instances into absence of CAD and 13 instances into presence of CAD. Similarly, under run-3, the IFS-DLRFP technique has classified a total of 363 instances into absence of CAD and 16 instances into presence of CAD. Additionally, under execution run-5, the IFS-DLRFP algorithm has categorized 364 instances into absence of CAD and 17 instances into presence of CAD.

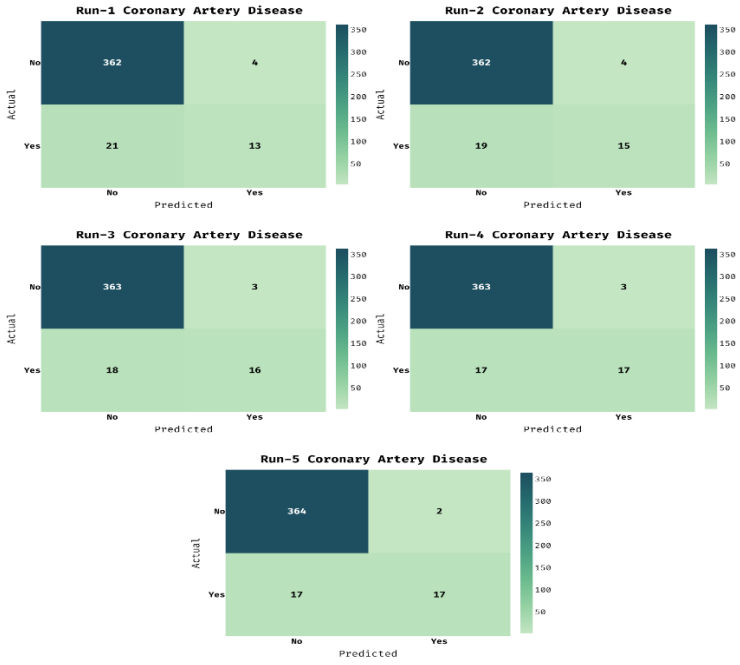

Fig. 5: Confusion matrix of IFS-DLRFP model on CAD classification

Table 2 offers the CAD classification outcomes analysis of the IFS-DLRFP technique under varying runs. The outcomes showcased that the IFS-DLRFP technique has attained efficient outcomes under every run. For instance, with run-1, the IFS-DLRFP approach has classified the CAD with the pre_n, rec_l, $spec_y$, acc_y, and F_{score} of 0.9452, 0.9891, 0.3824, 0.9375, and 0.9666 correspondingly.

Eventually, with run-3, the IFS-DLRFP manner has classified the CAD with the pre_n, rec_l, $spec_y$, acc_y, and F_{score} of 0.9528, 0.9918, 0.4706, 0.9475, and 0.9719 correspondingly. Meanwhile, with run-5, the IFS-DLRFP manner has

classified the CAD with the pre_n, rec_l, $spec_y$, acc_y, and F_{score} of 0.9554, 0.9945, 0.5000, 0.9525, and 0.9746 correspondingly.

Table 2: Result analysis of IFS-DLRFP model on CAD classification

No. of Runs	Precision	Recall	Specificity	Accuracy	F-Score
Run-1	0.9452	0.9891	0.3824	0.9375	0.9666
Run-2	0.9501	0.9891	0.4412	0.9425	0.9692
Run-3	0.9528	0.9918	0.4706	0.9475	0.9719
Run-4	0.9553	0.9918	0.5000	0.9500	0.9732
Run-5	0.9554	0.9945	0.5000	0.9525	0.9746
Average	**0.9518**	**0.9913**	**0.4588**	**0.9460**	**0.9711**

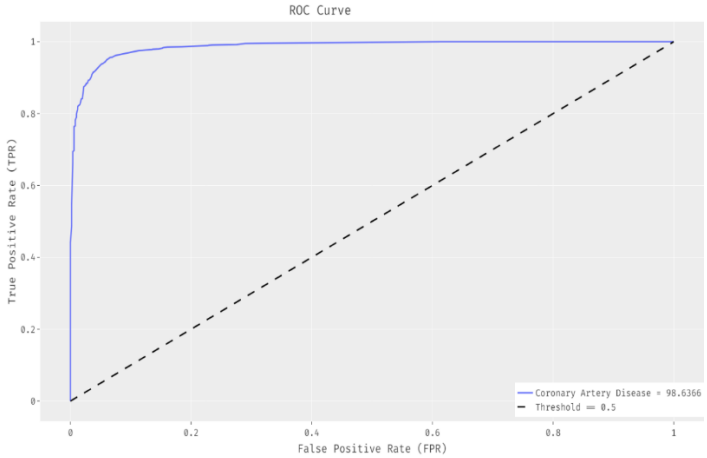

Fig. 6: ROC analysis of IFS-DLRFP model on CAD classification

Fig. 6 demonstrates the ROC analysis of the IFS-DLRFP algorithm on the classification of CAD. The figure is obvious that the IFS-DLRFP approach has resulted in a superior ROC of 98.6366%.

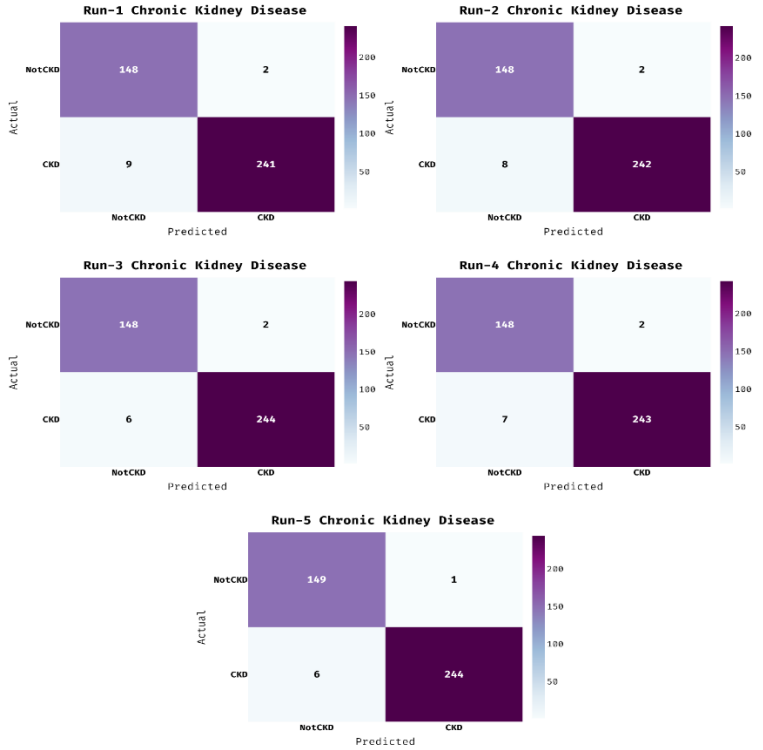

Fig. 7: Confusion matrix of IFS-DLRFP model on CKD classification

Fig. 7 portrays the confusion matrices produced by the IFS-DLRFP algorithm on the classification of CKD. On the execution run-1, the IFS-DLRFP manner has categorized 148 instances into absence of CKD and 241 instances into presence of CKD. In addition, under run-3, the IFS-DLRFP methodology has classified a total of 148 instances into absence of CKD and 244 instances into presence of CKD. Besides, under execution run-5, the IFS-DLRFP method has categorized 149 instances into absence of CKD and 244 instances into presence of CKD.

Table 3 provides the CKD classification outcomes analysis of the IFS-DLRFP manner under varying runs. The outcomes demonstrated that the IFS-DLRFP algorithm has reached effectual results under every run. For instance, with run-1, the IFS-DLRFP method has classified the CKD with the pre_n, rec_l, $spec_y$, acc_y, and F_{score} of 0.9427, 0.9867, 0.9640, 0.9725, and 0.9642 correspondingly. Eventually, with run-3, the IFS-DLRFP technique has classified the CKD with the pre_n, rec_l, $spec_y$, acc_y, and F_{score} of 0.9610, 0.9867, 0.9760, 0.9800, and 0.9737 respectively. Meanwhile, with run-5, the IFS-DLRFP system has classified the CKD with the pre_n, rec_l, $spec_y$, acc_y, and F_{score} of 0.9613, 0.9933, 0.9760, 0.9825, and 0.9770 correspondingly.

Table 3: Result analysis of IFS-DLRFP model on CKD classification

No. of Runs	Precision	Recall	Specificity	Accuracy	F-Score
Run-1	0.9427	0.9867	0.9640	0.9725	0.9642
Run-2	0.9487	0.9867	0.9680	0.9750	0.9673
Run-3	0.9610	0.9867	0.9760	0.9800	0.9737
Run-4	0.9548	0.9867	0.9720	0.9775	0.9705
Run-5	0.9613	0.9933	0.9760	0.9825	0.9770
Average	**0.9537**	**0.9880**	**0.9712**	**0.9775**	**0.9705**

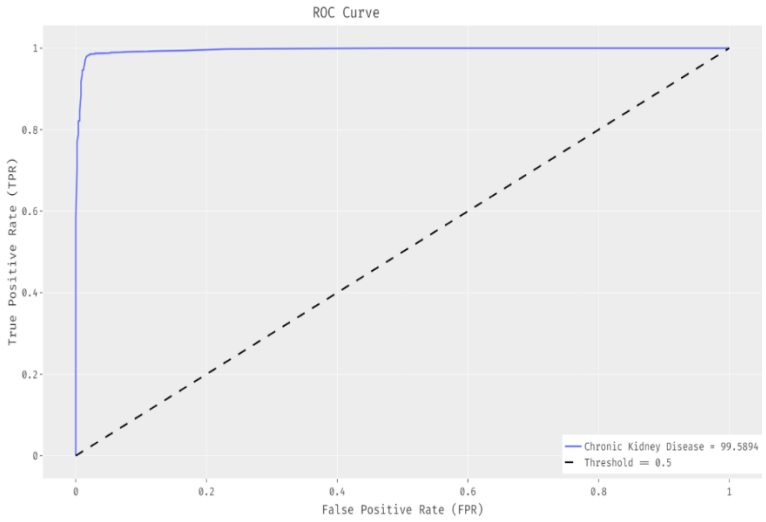

Fig. 8: ROC analysis of IFS-DLRFP model on CKD classification

Fig. 8 showcases the ROC analysis of the IFS-DLRFP manner on the classification of CKD. The figure exhibited that the IFS-DLRFP method has resulted in an increased ROC of 99.5894%.

Finally, a detailed comparative e result analysis of the IFS-DLRFP technique with existing ones takes place in Table 4 [22-25]. Fig. 9 showcases the accuracy analysis of the IFS-DLRFP technique with other existing techniques. The figure has shown that the Voted Perceptron, DT, LogitBoost, and LR techniques have accomplished lower accuracy values. At the same time, the AI-CHD, PSO, SVM, ACO, and MRODC techniques have tried to showcase somewhat moderately closer accuracy values. However, the presented IFS-DLRFP technique has resulted in higher accuracy of 0.9383.

Table 4: Result Analysis of Proposed IFS-DLRFP Model interms of Various Measures

Methods	**Precision**	**Recall**	**Accuracy**	**F-score**
IFS-DLRFP	0.9501	0.9553	0.9383	0.9522
AI-CHD	0.8000	0.2041	0.8088	0.5618
ACO	0.8734	0.8888	0.8750	0.9056
PSO	0.8624	0.8800	0.8500	0.8800
MRODC	0.9180	0.9089	0.8867	0.9134
Logistic Regression	0.8800	0.7927	0.7721	0.8341
Voted Perceptron	0.9240	0.6804	0.6679	0.7837
LogitBoost	0.8460	0.7761	0.7408	0.8095
Decision Tree	0.8140	0.7902	0.7382	0.8019
SVM Model	0.8686	0.8710	0.8687	0.8822

Fig. 10 illustrates the outcome analysis of the IFS-DLRFP manner with other recent approaches. The figure demonstrated that the Voted Perceptron, DT, LogitBoost, and LR manners have accomplished minimum precision, recall, and F-score values. Simultaneously, the AI-CHD, PSO, SVM, ACO, and MRODC methods have tried to

illustrate slightly moderately closer precision, recall, and F-score values. Eventually, the projected IFS-DLRFP methodology has resulted in superior precision, recall, and F-score of 0.9501, 0.9553, and 0.9522.

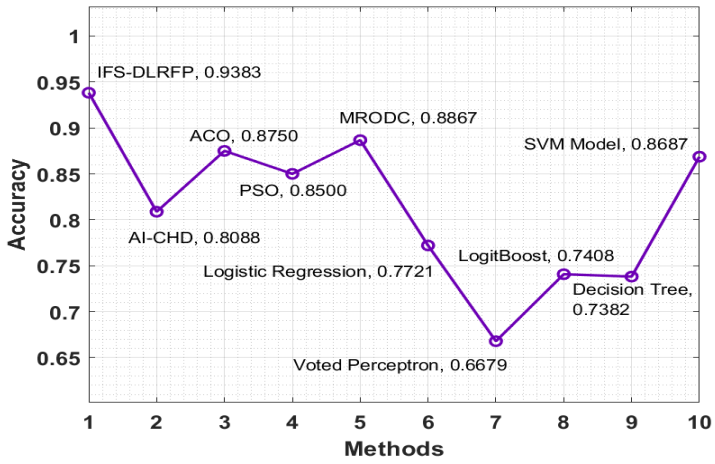

Fig. 9: Accuracy a nalysis of IFS-DLRFP model with existing techniques

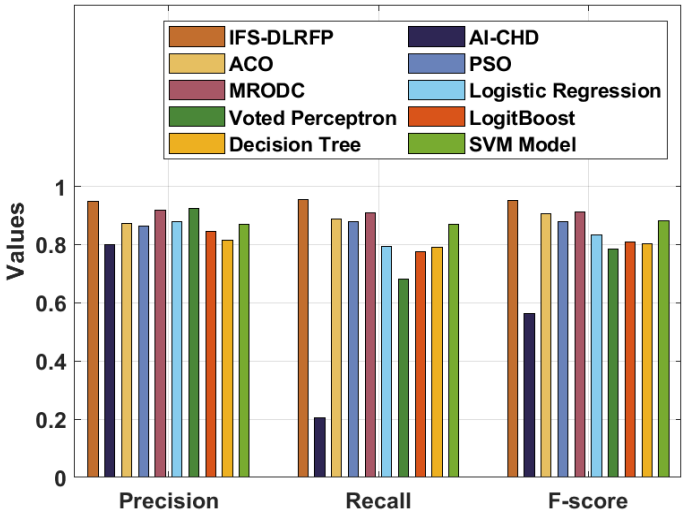

Fig. 10: Comparative analysis of IFS-DLRFP model with different measures

From the detailed results analysis, it is obvious that the IFS-DLRFP technique has found to be a proficient tool for disease diagnosis in real time environment. It is due to the inclusion of FFOA based FS and hyper parameter optimization.

5.0 CONCLUSION

In this study, a novel IFS-DLRFP approach is developed for determining the risk factors of CKD and CKD among T2DM patients. The proposed IFS-DLRFP technique encompasses four processes namely pre-processing, FFOA based FS, GRU based classification, and FF based hyper parameter optimization. The FF-GRU technique carries out the hyper parameter tuning process by the use of FF technique. The design of FS and hyper parameter tuning processes results in improved classification performance. For assessing the effectual results of the IFS-DLRFP technique, a comprehensive experimental analysis is carried out on benchmark dataset and the simulation results stated the improved performance of the IFS-DLRFP technique compared to other techniques. Therefore, the IFS-DLRFP technique can be employed as a proficient tool for effective diagnosis in real time. In future, feature reduction and hybrid DL models can be applied to furthermore improve the predictive performance.

REFERENCES

[1] I. Martín-Timón, C.Sevillano-Collantes, C., Segura-Galindo, A. and del Cañizo-Gómez, F.J., "Type 2 diabetes and cardiovascular disease: have all risk factors the same strength?" *World journal of diabetes*, 5(4), 2014, p.444.

[2] JM.Evans, J.Wang and AD.Morris. "Comparison of cardiovascular risk between patients with type 2 diabetes and those who had had a myocardial infarction: cross sectional and cohort studies". *BMJ* 2002; 324: 939-942.

[3] FJ.delCañizo-Gómez and MN.Moreira-Andrés. "Cardiovascular risk factors in patients with type 2 diabetes. Do we follow the guidelines?" *Diabetes Res Clin Pract* 2004; 65: 125-133.

[4] Matsushita, K., Ballew, S.H. and Coresh, J., "Cardiovascular Risk Prediction in People with CKD". *Current opinion in nephrology and hypertension*, 25(6), 2016:p.518.

[5] Bibbins-Domingo K.., "Aspirin Use for the Primary Prevention of Cardiovascular Disease and Colorectal Cancer: U.S. Preventive Services Task Force Recommendation Statement Aspirin Use for the Primary Prevention of CVD and CRC". *Annals of Internal Medicine*. 2016.

[6] Matsushita K, van der Velde M, Astor BC, Woodward M, Levey AS, de Jong PE, Coresh J, Gansevoort RT. "Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis". *Lancet*. 2010; 375:2073–81.

[7] Krishnamurthy, S., Ks, K., Dovgan, E., Luštrek, M., Gradišek Piletič, B., Srinivasan, K., Li, Y.C.J., Gradišek, A. and Syed-Abdul, S., May. "Machine learning prediction models for chronic kidney disease using national health insurance claim data in Taiwan". *In Healthcare Multidisciplinary Digital Publishing Institute.* Vol. (9), No. 5,2021: p. 546.

[8] A.Callahan, N.H.Shah, "Machine Learning in Healthcare. In Key Advances in Clinical Informatics", *Elsevier: Amsterdam, The Netherlands*, 2017; pp. 279–291.

[9] A.J.Collins, J.A.Vassalotti, C.Wang, S.Li, D.T.Gilbertson. J.Liu, A.N.Foley, S.-C.Chen, T.J.Arneson, "Who Should Be Targeted for CKD Screening? Impact of Diabetes, Hypertension, and Cardiovascular Disease". *Am. J. Kidney Dis*. 2009, 53, S71–S77.

[10] M.E.Hossain, S. Uddin and A.Khan, "Network analytics and machine learning for predictive risk modelling of cardiovascular disease in patients with type 2 diabetes". *Expert Systems with Applications*, 164, 2021. p.113918.

[11] Jamthikar, A.D., Gupta, D., Johri, A.M., Mantella, L.E., Saba, L., Kolluri, R., Sharma, A.M., Viswanathan, V., Nicolaides, A. and Suri, J.S., *Low-cost office-based cardiovascular risk stratification using machine learning and focused carotid ultrasound in an Asian-Indian cohort*. Journal of Medical Systems, 44(12), 2020:pp.1-15.

[12] J.K.Kim, and S.Kang, "Neural network-based coronary heart diseases risk prediction using feature correlation analysis". *Journal of healthcare engineering*, 2017.

[13] E.M.Senan, M.H. Al-Adhaileh, F.W. Alsaade, T.H. Aldhyani, A.A. Alqarni, N.Alsharif, M.I.Uddin, A.H. Alahmadi, M.E. Jadhav, and M.Y.Alzahrani, "Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques". *Journal of Healthcare Engineering*, 2021.

[14] Song, X., Waitman, L.R., Alan, S.L., Robbins, D.C., Hu, Y. and Liu, M., "Longitudinal risk prediction of chronic kidney disease in diabetic patients using a temporal-enhanced gradient boosting machine: retrospective cohort study". *JMIR medical informatics*, *8*(1), 2020: p.e15510.

[15] Z.Segal, D. Kalifa, K. Radinsky, B. Ehrenberg, G. Elad, G. Maor, M.Lewis, M.Tibi, L. Korn, and G. Koren, "Machine learning algorithm for early detection of end-stage renal disease". *BMC Nephrology*, 21(1), 2020.pp.1-10.

[16] K.B.Chimwayi, N.Haris, R.D.Caytiles, and N.C.S.Iyengar, "Risk level prediction of chronic kidney disease using neuro-fuzzy and hierarchical clustering algorithm (s).", *International Journal of Multimedia and Ubiquitous Engineering*, 12(8),2017, 23-36.

[17] Y.Li, and M.Han, "Improved fruit fly algorithm on structural optimization. Brain informatics*", Brain Informatics*, 7(1), 2020:pp.1-13.

[18] Y.Hou, J.Li, H.Yu and Z.Li, "A novel binary improved fruit fly algorithm for feature selection". *IEEE Access*, *7*, 2019.pp.81177-81194.

[19] Bi, L., Hu, G., Raza, M.M., Kandel, Y., Leandro, L. and Mueller, D., "A Gated Recurrent Unit (GRU)-Based Model for Early Detection of Soybean Sudden Death Syndrome through Time-Series Satellite Imagery". *Remote Sensing*, 12(21), 2020:p.3621.

[20] A.A. Ewees, M.A. Al-qaness and M. Abd Elaziz "Enhanced salp swarm algorithm based on firefly algorithm for unrelated parallel machine scheduling with setup times". *Applied Mathematical Modelling*, 94, 2021:pp.285-305.

[21] https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease

[22] R. Fan, N. Zhang, L. Yang, J. Ke, D. Zhao and Q.Cui, "AI-based prediction for the risk of coronary heart disease among patients with type 2 diabetes mellitus". *Scientific reports* 2020, 10(1), pp.1-8.

[23] M.Elhoseny, K. Shankar and J.Uthayakumar, "Intelligent diagnostic prediction and classification system for chronic kidney disease". *Scientific reports*, 9(1), 2019. pp.1-14.

[24] R.T. Selvi and I. Muthulakshmi, Modelling the map reduce based optimal gradient boosted tree classification algorithm for diabetes mellitus diagnosis system. *Journal of Ambient Intelligence and Humanized Computing*, 2020. pp.1-14.

[25] M.S. Amin, Y.K. Chiam and K.D.Varathan, "Identification of significant features and data mining techniques in predicting heart disease". "*Telematics and Informatics", 36*, 2019: pp.82-93.