Rev. Roum. Sci. Techn.– Électrotechn. et Énerg.
Vol. **68**, *1*, pp. 113–118, Bucarest, 2023

Génie biomédical

A DISEASE PREDICTION MODEL USING SPOTTED HYENA SEARCH OPTIMIZATION AND BIDIRECTIONAL LONG SHORT-TERM MEMORY

AYSHWARYA, LAKSHMI

Keywords: Machine learning; Deep learning; Optimization; Spotted hyena optimization algorithm; Cuckoo search optimization algorithm; Bidirectional long short-term memory; Prediction; Classification and fuzzy temporal.

Early disease prediction is the best solution for any deadly disease in this fast internet world. In this direction, machine learning and deep learning techniques are applied in this fast world to predict diseases early and update the disease level through the Internet of Things techniques. For this purpose, this paper proposes a disease prediction system that uses the newly proposed fuzzy temporal correlation aware classifier and the auto-encoded bidirectional long short-term memory (FTC-Bi-LSTM) to predict the diseases such as heart, cancer, and diabetes. Moreover, this paper proposes a hybrid optimization algorithm called spotted hyena and cuckoo search optimization algorithm (SHCSA) to select the contributed features used to enrich the prediction accuracy. The standard benchmark UCI repository dataset is used to conduct the various experiments and obtain better performance in terms of precision, recall, f-measure, and prediction accuracy. The result is validated by considering the hospital patient data as input and proving the performance.

1. INTRODUCTION

Recently, the world has been severely affected by various deadly diseases, including cancer, diabetes, and heart disease, due to the lack of awareness about the disease level and seriousness and food habits. Due to work and other pressures, heart disease affects too many people, irrespective of age group.

Cancer disease affected number is also gradually increasing in this fast world practicing irregular and unhealthy food consumption. Among them, diabetes disease is generally identified as diabetes by physicians. It is clearly stated that the human body cannot do blood glucose or blood sugar level [1]. This disease affects most people around the globe, and it can be categorized as type-1 diabetes and type-2 diabetes [2]. Heart diseases are very dangerous and damage the heart due to depression, irregular food habits, cholesterols, lack of physical activity, and stress, even though it is destructive and cause dangerous problems, especially in the nerves, kidneys, and eye blood vessels [3]. According to the World Health Organization (WHO) report released in 2014, globally, the number of adult diabetes patients increased by double the total cases in 1980, from 4.7 % to 8.5 %. Much research has been done on diabetes disease prediction by applying machine learning (ML) techniques. ML has demonstrated its ability to handle disease datasets efficiently and dealt with many features while designing a novel disease prediction model. The supervised ML methods focus on finding the relevant terms in the form of the relevant features and terms.

Data pre-processing is the most important task for enhancing disease prediction performance. Here, feature selection, extraction, and optimization play a vital role and are widely used by many researchers in many applications. The major benefits of the feature reduction techniques are related to various manners in the data analysis process, including understanding the data, data visualization, and prediction accuracy—two types of feature selection methods, such as filter and wrapper methods. In addition, various optimization techniques are incorporated in the feature selection process for identifying and selecting the contributed features, which are used to enhance the prediction result on various data mining algorithms, machine learning, and deep learning algorithms.

With the help of the feature selection process, the ML algorithms are achieving better results on various datasets such as medical datasets, network traffic datasets, e-learning datasets, amazon datasets, and weather datasets. Even though ML cannot produce the expected accurate prediction result for various datasets due to the presence of huge data, to overcome these all, the advancement of ML algorithms is introduced as Deep Learning algorithms such as deep belief network (DBN), convolutional neural network (CNN) and long-short term Memory (LSTM) for enhancing the performance in the training process and enhances the accuracy. However, optimization techniques are also used to improve prediction accuracy by reducing the number of most important features and contributing more to making effective decisions on the many datasets by deep learning algorithms. For this purpose, this paper proposes a disease prediction system combining an existing feature optimization technique and the effective deep learning technique.

The Research objective of this work is to predict diseases early by analyzing the patient records using a disease prediction system and to safeguard the patients.

The main contributions are listed below:

1. To propose a new disease prediction model to predict heart, cancer, and diabetes effectively.
2. To propose a new feature optimization technique named spotted hyena and cuckoo search optimization algorithm (SHCSA) for performing an effective feature optimization process that enhances prediction accuracy.
3. To apply the K-means clustering algorithm for grouping the patient records concerning the feature levels of the various datasets.
4. To apply the new proposed fuzzy temporal correlation aware classifier with auto-encoded BLSTM (FTC-Bi-LSTM) for effectively predicting the diseases and their levels like very high, high, medium, and low.

The rest of this paper is organized as follows: The related works with merits and demerits are described in section 2. The proposed system architecture is shown in section 3. Section 4 demonstrates the workflow of the proposed model with algorithms and explanations. Section 5 shows the performance of the proposed model. Section 6 concludes the work with future directions.

Department of Computer Science and Engineering, University College of Engineering Panruti, Tamilnadu, India.
Email: ayshresearch@gmail.com

2. RELATED WORKS

Different researchers have done many research works. A new fuzzy temporal cognitive map (FTCM) for predicting diabetes, cancer, and heart was proposed in the paper [4]. The FTCM applied neural networks and temporal constraints for making effective decisions. A new neural classifier called Fuzzy Temporal Min-Max Neural network with Particle Swarm optimization model to predict the diseases such as diabetes, cancer, and heart diseases effectively was developed in the paper [4]. They have achieved better performance in terms of prediction accuracy than the existing classifiers.

A new rule extraction method that combines the sampling selection methods for achieving high accuracy on Pima Indian dataset they have achieved around 84 % accuracy, was developed in the paper [5]. A new hybrid method was developed, incorporating the neural network and SVM for categorizing the inter-class and intra-class transitions for predicting the number and the range of beta regions and presented by the author [6]. Their method uses a sliding window-aware feature selection technique for extracting the suitable features that help enhance the classifier's performance in the training process. An extensive review of the various machine learning and data mining algorithms in the past for predicting diabetes achieved better accuracy than the existing classifiers available in the literature [7].

The author proposed new fuzzy rule-based reinforcement learning-based evolutionary system for diagnosing diabetes and presented it in the paper [8]. They have incorporated the rule base and the optimization technique to enhance the performance of their system. They have developed a new rule base that has learned about the initial rules and numeric values and removed the redundant rules. They have applied suitable rules with interpretability for performing the pruning process and used genetic algorithms and reinforcement learning to achieve better accuracy. They have proved their system is better than other models by achieving the highest prediction accuracy on diabetes datasets. The author of the paper [9] considered the ten thousand records that were collected from 36 different hospitals in the years between 2008 and 2016. They have performed the prediction process using well-known classifiers such as support vector machine (SVM), neural classifier, logistic regression, and decision trees for identifying the disease-affected patient records.

The author of the paper [10] proposed a new hybrid method to extract the rules using an SVM classifier. Moreover, they have incorporated a new feature selection method to select the significantly grouped features from the benchmark dataset. They have applied XG Boost for converting the SVM black box model considering live data as input. They have proved that their model performed better than the existing models by considering other standard and live datasets. The author of the paper [11] proposed a new k-means clustering-based outlier method for detecting outliers. It selects the most important features by applying meta-heuristic methods to finalize effective features that can predict diseases. The standard classifier SVM is used for performing classification.

Moreover, they conducted various experiments and did cross-validation to ensure the accuracy of classification. The author of the paper [12] proposed a new cancer disease prediction system incorporating fuzzy temporal logic for predicting diseases such as cancer, heart disease, and diabetes. They have achieved good prediction accuracy than the existing systems.

In the paper [13], the author developed a new ensemble classifier that combines the roughest filter-based optimal subset selection and classification method. They have proposed a new intersection method for reducing the processing time.

The author proposed a new stacking-aware evolutionary ensemble method to predict type-2 diabetes in the paper [15]. Their method identified the missing values and outliers for identifying the mean values. Moreover, a multi-objective optimization method enhances classification accuracy and reduces complexity. Their method achieved 84 % prediction accuracy.

The authors have developed a category boosting (CatBoost) ML algorithm in [21] to predict the various stages of breast cancer that facilitate early diagnosis. The PSO toll (PSOT) resolved the online parameters estimation of permanent magnet synchronous machines, increasing the speed [22].

Many researchers for predicting many diseases have proposed various models. Even though there is no system available for fulfilling the current users' expectations to predict the accurate disease stage of *Diabetes Mellitus*, and not achieved enough prediction accuracy. For this purpose, this paper proposes a new disease prediction model incorporating a new hybrid feature optimization technique, clustering method, and deep learning algorithm for performing an effective prediction process.

3. SYSTEM ARCHITECTURE

The proposed model is demonstrated in Fig. 1, explaining various functionalities using ten components: a user interaction module, database, rule base, decision manager, rule manager, data pre-processing module, and prediction module.

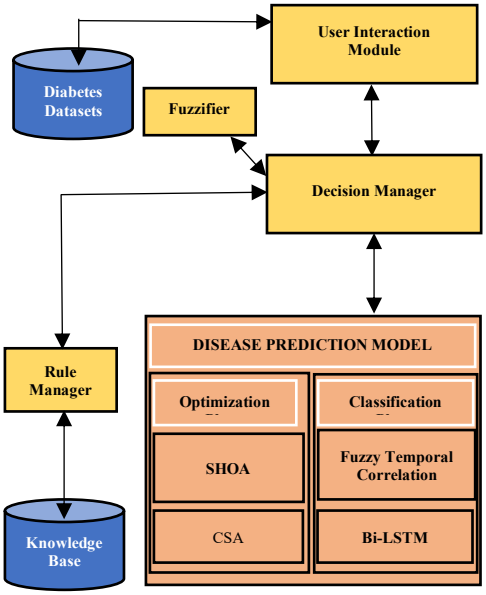

Fig. 1 – System architecture

The user interaction module collects the necessary data from diabetes datasets. The user sent the request to the decision manager. The decision manager retrieves the relevant data from the diabetes dataset through the user interaction module. The decision manager processes the data by using the prediction module. The prediction module contains two different phases as optimization phase and the classification phase. In the optimization phase, feature

optimization is performed by applying a newly proposed hybrid optimizer that combines SHOA and cuckoo search algorithm (SHCSA).

In the classification phase, the optimized feature sets of records are considered input for the classification module, and it performs the classification process using the deep learning algorithm. Before that, the relevant patient records can be summarized by applying the existing clustering algorithm called K-means clustering to group the datasets effectively. Moreover, it classifies using a fuzzy temporal correlation and the Auto-encoder aware Bi-LSTM. The decision manager also performs the fuzzification with the help of a fuzzifier. The decision manager makes an effective decision on patient records by applying the necessary facts and rules in the knowledge base.

4. **PROPOSED MODEL**

This section explains in detail the proposed disease system that uses the newly proposed feature optimization algorithm SHCSA which combines SHOA CSA [20] for selecting the most contributed features. Moreover, this work proposes a fuzzy temporal correlation-based ensemble classifier for effective classification. This fuzzy temporal correlation-based classifier incorporated with autoencoder-based bidirectional long short-term memory (Bi-LSTM) is also used for effective classification in this work. This section explains in detail the feature optimization process first, and it explains the classification process.

4.1. FEATURE OPTIMIZATION

This subsection explains in detail the newly proposed feature optimization method SHCSA used for an effective feature selection process. First, it explains the background of the SHOA. Finally, it explains the newly proposed SHCSA in detail with the necessary steps for performing effective feature selection and optimization processes in this work.

4.2. SPOTTED HYENA OPTIMIZATION

Generally, spotted hyenas are considered huge carnivorous canines living in different kinds of dry and open environments. The different sizes, like medium-size and large-size herbivores, namely zebras, impala, and wildebeests, are preyed by spotted hyenas' prey. Spotted hyenas are social animals with intelligence, and they can be identified by relatives applying the various senses. Moreover, they are ranked in race according to the relationships between them. In nature, the spotted hyenas can achieve a success rate in group hunting. It contains five steps: encircling prey, hunting, exploiting, and exploring. Here, the encircling process selects the best candidate nearer to the target prey. Moreover, update the search agent locations with definitions. In the hunting process, the spotted hyenas hunt in packs, rely on friends of trusted networks, and can spot prey. For defining the behavior of the spotted hyena, consider the best agent: the optimal location of the targeted prey. The location of the targeted prey is updated based on the best solution. In the exploitation process, the vector value is reduced to attack the prey. The spotted hyena group attacks the prey and changes the search agent's location based on the best agent solutions. This SHO permits the search agents to update the locations and the relevant attacks of the prey. In the exploration process, the SH's find keys based on the SH's group. Moreover, they are in different places to find prey of attacks. It permits the global search that makes the possible exploration processes. In addition, the random weights are assigned to every prey and take precedence for demonstrating the randomness that avoids local optimization and global search processes. Especially, local optimization is avoided in the final iteration process and terminates when satisfied the given condition.

PROPOSED SHCSA

This section explains the proposed SHCSA in detail with the necessary steps. This new optimization method is used to select the useful features capable of enhancing the classifier's performance.

Algorithm: Spotted Hyena and Personalized Cuckoo Search Optimizer (SHPCSO)
Input: Datasets
Output: Optimised feature set
Phase 1:
Step 1: Perform the encircling prey (optimal and useful feature)
Step 2: Perform the hunting process to identify the useful features
Step 3: Perform the exploitation process on the identified features
Step 4: Perform the exploration task on the features
Phase 2:

```
Cuckoo-Search (Pa, a, λ)
{
 Nests = n;
 While (i<=n)
  {
     Read Cuckoo (say i)
     Calculate the Fitness values for the feature Fi
     Choose a nest from N randomly
     IF Fi is better than Fj THEN
         Fi=Fj
         Pa is declared as abandoned
         NA = Pa
         Rank (nests)
         Best nest (n)
   }
  Return (nest)
}
```

The proposed SHPCSO combines SHOA and the proposed enhanced personal cuckoo search algorithm (EpBestCSA) for optimizing the features to categorize the contents. Phase 1, called the enhanced SHOA, performs the initial level feature selection. Phase 2 covers the steps of the EpBestCSA, and it is applied for performing the second-level optimization in the proposed optimizer. The optimal featured record sets are grouped by applying the K-means clustering algorithm for further process.

4.4. CLASSIFICATION

This section explains the newly proposed fuzzy temporal correlation-based ensemble classifier and the necessary background information. First, it explains the fuzzy correlation with temporal constraints. Then, it explains the fuzzy temporal cognitive map and Autoencoder Bi-LSTM consecutively.

4.4.1. FUZZY CORRELATION WITH TEMPORAL CONSTRAINTS

In the proposed prediction system, the fuzzy correlation values and temporal constraints are also considered to make a final decision on the disease dataset according to work [8]. The coefficient value of fuzzy correlation is considered

in the classification process to achieve better prediction accuracy. The methodology is explained as follows:

First, the $Xn = \{x1, x2, ...\}$ and $Yn = \{y1, y2, ...yn\}$ are the independent observation, then the correlation coefficient relation is derived from eq. (4).

$$r(t1,t2 = ni = 1(Xi - \overline{\overline{X}})(Yi - \overline{\overline{Y}}) \quad (4)$$

$$ni = 1\ (Xi - \overline{\overline{X}})2 \sum n(Yi - \overline{\overline{Y}})2$$

Apply the fuzzification process with the variables *Xn* and *Yn* by applying the triangular membership function and defining the non-membership in eq. (5) and (6).

$$(x) < t1,t2 >= x - pifp \le x \le q$$

$$q - pr - xifq \le x \le (t1,t2) \quad (5)$$

$$r - q10 \quad \text{if x<pandx> (t1,t2)}$$

$$\text{and}\,(x)\text{<t1,t2>=}\{q - xq - p * x - qr * - qifp * \le x \le q$$

$$\text{if q}\ ifq \le x \le r * \quad (6)$$

$$1\ \ ifx < p * andx > r *$$

where p*<p<q<r<r* now, to perform fuzzification for the value *r,* and it calculates the alpha-beta cut values of y as:

$$(t1,t2) =$$

$$= [\{(t1,t2), (t1,t2) \ge \alpha\}, \{(t1,t2), (t1,t2) \ge \alpha\}] \quad (7)$$

$$(t1,t2) =$$

$$= [\{(t1,t2), (t1,t2) \ge \beta\}, \{(t1, t2), (t1, t2) \ge \beta\}] \quad (8)$$

And from eq. (3) and (4), this work has

$$r\alpha(t1,t2) = [\min\{r(t1,t2) \in [-1,1], r(t1,t2) \ge \alpha\},$$

$$\max\{\gamma(t1,t2), \gamma(t1,t2) \ge \alpha\}] \quad (9)$$

$$r\beta(t1,t2) = [\min\{\gamma(t1,t2) \in [-1,1], \gamma(t1,t2) \ge \beta\},$$

$$\max\{r(t1,t2), r(t1,t2) \ge \beta\}] \quad (10)$$

From equ. (5) to (7), it is found that to calculate the correlation value (γ ') is easy below:

$$r'(t1,t2) = (r\ \alpha\ (t1,t2) + r\ \beta\ (t1,t2) - \gamma)\ \text{defu} +$$

$$(1 - r\ \alpha\ (t1,t2) - r\ \beta\ (t1,t2) + \gamma)\ \text{defv} \quad (11)$$

where $\gamma = (r\ \alpha, r\ \beta)$ indicates the contradictory methodology.

The newly proposed ensemble classifier is a meta-classifier to merge two or more ML algorithms to predict the disease using majority voting. A soft voting classification method applies two types of voting methodologies such as hard voting and soft voting. In the hard voting technique, the final decision is made on predictive analysis by considering the majority voting that aggregates and chooses the class which repeatedly appears from the models. The base models must incorporate the Predict_proba technique in the soft voting methodology. The voting classification method presents better results than the base models, which combine the various models. The Fuzzy Temporal Cognitive Map and the Auto encoding Bi-LSTM have been ensembled in the proposed model. The predict_proba feature is applied for every targeted feature and mixes the training data and the points. These data points are passed to the classifiers FTCM and Bi-LSTM. Every model calculates the voting aggregation value and also finds the yields the final disease prediction process. The steps of the proposed fuzzy temporal aware soft voting classifier are below:

Procedure Replace_data (dd) // dd represents the diabetes data
Return dd ["pregnancy", "insulin","BMI"].replace ('0', median()) Procedure DS (dd) // DS represents the data split
TRD, TSD = split (DF, label) // DF -Diabetic features, TRD – Training data, TSD – Testing data
Return (TRD, TSD) Voting =" soft"
M1= FTCM (TRD, TRL, TSD) // TRL – Training Labels M2= SVM (TRD, TRL, TSD)
Procedure EM (TRD, TRL, TSD)
SVC = concatenate (M1, M2) // SVC – Soft Voting Classifier SVC.fit (TRD, TRL)
Predictions = SVC.predict(TSD)

The necessary rules were developed for predicting diseases such as heart disease and cancer according to the diabetes disease prediction process.

4.4.2. CLASSIFICATION USING AUTOENCODING BI-LSTM

This section explains the workflow of the Bi-LSTM in detail by demonstrating the necessary formulae used in this work for performing air quality prediction. This section explains the standard deep learning algorithms, namely LSTM and Bi-LSTM, used for predicting air quality according to work [15,16]. The LSTM procedure is discussed with the consideration of the cell and the three different gates helpful in determining the cell states. Here, the forget gate fg_i is one of the layers in a neural network that can be shown as eq. (12).

$$f_t = \sigma(W_f[h_{i-1}, X_t] + b_f], \quad (12)$$

where σ represents the sigmoid function, h_{i-1} indicates the previous output, X_t means the current input, W_f represents the weights, and b_f indicates the biases. Here, the input gate is considered another neural network and can create a forget gate (ft) like the formula in eq. (2)

$$i_t = \sigma(W_i[h_{i-1}, X_t] + b_i), \quad (13)$$

where the weights and biases hold different values.

The candidate value is represented by using the variable C_t along with the values of h_{t-1} and X_t in the equation (13).

$$C_t = \tanh(W_C[h_{i-1}, X_t] + b_C), \quad (14)$$

Next, change the cell state $C_{t,\ multiply}$ the historical state $C_{t-1,}$ and forget the gate. Moreover, multiply the input gate along with candidate state C_t and add them. Finally, it provides the structure given in equation (15).

$$C_t = f_t * C_{t-1} + i_t * C_t \quad (15)$$

Still, the cell state is updated. Finally, the output of the LSTM, according to the state of the cell that, is the combination of the cell state and output gate.

$$o_t = \sigma(W_o[h_{i-1}, X_t] + b_o),$$

$$h_t = o_t * \tanh(C_t). \quad (16)$$

The necessary features are trained by applying the backpropagation through time method that is shortly called BPTT, which has a backpropagation variant. This work has been implemented using tensor flow and applies an

optimization method named Adam optimizer to learn the dataset's features.

5. RESULTS AND DISCUSSION

This section explains the dataset, evaluation parameters, and experimental and comparative results. First, it explains the datasets used in this work in detail.

A. DATASETS

This subsection explains the standard benchmark dataset and the live patient records that are considered in this work to prove the effectiveness and efficiency of the proposed disease prediction model. From the paper [17], the standard benchmark dataset called UCI machine learning repository dataset to validate the prediction result of the proposed model in terms of evaluation metrics. The experiments were done by considering 11 features selected from heart disease datasets, such as chest pain, sex, blood pressure, cholesterol, blood sugar, EKG, maximum heart rate, exercise, old peak, thallium, and age are used to predict heart diseases. The blood glucose level is considered one of the important features in predicting diabetic disease. If the blood glucose level is above 100 mg/dl, the patient is affected with diabetes. At the same time, if the heart rate is above 100, then there is also possible to affect by heart disease. It especially uses the UCI machine learning repository Pima Indian dataset for conducting experiments to predict diabetes mellitus with classifications such as type-1 and type-2. The dataset has been divided into testing and training datasets with 30 % and 70 %, respectively.

B. EXPERIMENTAL SETUP

The experiments have been done to prove the effectiveness and efficiency of the proposed disease prediction model, which incorporates IoT devices, the relevant source code, and Raspberry PI software. Moreover, the disease prediction model incorporates IoT devices, hardware devices, microcontrollers, and LoRa communicators to transmit the data to the prediction model. Here, the Omron HeartGuid-bp8000m is used for measuring blood pressure, and the heart monitor board is also used to measure the heart rate.

C. PERFORMANCE EVALUATION PARAMETERS

The proposed model evaluates based on accuracy, precision, recall, and f1 score. Here, the true positive (TP), true negative (TN), false negatives (FN), and false positives (FP) are considered for calculating the prediction accuracy by using the formulae given in the equations:

Acc

$$\text{Accuracy} = TP+TN/TN+TP+FP+FN \quad (17)$$

$$\text{Precision} = TP/TP+FP \quad (18)$$

$$\text{Recall} = TP/TP+FN \quad (19)$$

$$F1 - \text{Score} = 2 \text{ x Precision x Recall/Precision+Recall} \quad (20)$$

This work uses these parameters to evaluate the performance of the proposed disease prediction model effectively according to the precision, recall, f1-score, and accuracy calculated using the values of true positives, true negatives, false positives, and false negatives.

D. EXPERIMENTAL RESULTS

The disease prediction process can be done by considering the evaluation parameters such as precision, recall, and f1-score by considering all the features of the datasets.

Table 2 demonstrates the prediction analysis on datasets with reduced features by considering the precision, recall, and f1-score. Five different experiments, such as E1, E2, E3, E4, and E5, have been conducted to prove the effectiveness of the proposed disease prediction model. Here, a different set of records were considered for conducting the five experiments.

Table 2

Prediction result analysis on datasets with selected features

Exp.	Precision	Recall	F1-Score
E1	97.3	96.9	97.2
E2	96.3	96.4	96.4
E3	96.2	96.2	96.3
E4	97.9	97.7	97.6
E5	98.7	98.8	98.4

Table 2 proves the effectiveness of the proposed disease prediction model on various sets of records considered in five different experiments. The performance of the proposed prediction model is better in all the evaluation metrics, such as precision, recall, and f1-score, when it compared with the full-featured dataset. The enhancement is the incorporation of the proposed feature optimization method and fuzzy temporal correlation-based Auto-encoding Bi-LSTM.

Table 2 demonstrates the comparative performance analysis between the proposed model and the existing prediction systems.

Table 2

Comparative analysis

Model	E1	E2	E3	E4	E5
Ganapathy et al [18]	89.52	89.35	89.23	89.21	88.47
Renji et al [19]	91.52	91.35	91.23	92.21	91.47
Kanimozhi et al [12]	93.52	93.35	92.98	94.22	93.42
Elizabeth and Shabnam [23]	96.45	96.32	95.91	96.96	96.17
Proposed System	97.34	97.45	96.92	97.69	97.51

Table 2 proves the effectiveness of the proposed model when compared with the existing disease prediction models developed by the authors [12,18,19,23] in terms of prediction accuracy by obtaining 97.38 % overall accuracy. The performance improvement is applying an effective feature optimization technique incorporating the spotted hyena optimization algorithm, the cuckoo search optimization algorithm, and the newly proposed classifier that applies new fuzzy temporal correlation and the existing auto-encoding-based Bi-LSTM.

6. CONCLUSION AND FUTURE WORKS

A disease prediction system has been developed to predict the diseases such as diabetes, heart, and cancer. It uses the proposed SHCSOA to identify and select the most contributed features. Moreover, a newly developed fuzzy temporal correlation-based ensemble classifier is incorporated for effective classification. In addition, the proposed fuzzy temporal correlation and autoencoder-based Bi-LSTM are also used for enhancing prediction accuracy. The proposed model achieved 97.38 % overall prediction accuracy on the cancer, heart, and diabetic disease prediction process. This work achieves better prediction accuracy than

other works. Future works can be done in this direction by introducing a lightweight optimizer and classifier for making an effective decision on disease datasets.

Received on 24 January 2021

REFERENCES

1. R. Joshi, M. Alehegn, *Analysis and prediction of diabetes diseases using machine learning algorithm: Ensemble approach*, Int. Res. J Engg. Tech., **4**, *10*, pp. 426–435 (2017).
2. N.J.F. Ndisang, A. Vannacci, S. Rastogi, *Insulin Resistance, Type 1 and Type 2 Diabetes, and Related Complications*, J Diab. Res., 2017, id:1478294 (2017).
3. H.P. Himsworth and R. B. Kerr, *Insulin-sensitive and Insulin-insensitive types of diabetes mellitus*, Clinical Sciences, **4**, pp. 119–152 (1939).
4. R. Sethukkarasi, S. Ganapathy, P. Yogesh, A. Kannan, *An intelligent neuro fuzzy temporal knowledge representation model for mining temporal patterns*, Journal of Intelligent & Fuzzy Systems, **26**, *3*, pp. 1167–1178 (2014).
5. Y. Hayashi, S. Yukita, *Rule extraction using Recursive-Rule extraction algorithm with J48 graft combined with sampling selection techniques for the diagnosis of type 2 diabetes mellitus in the Pima Indian dataset*, Informatics in Medicine Unlocked, **2**, pp. 92–104 (2016).
6. H. Kazemian, S.A. Yusuf, K. White, C.M. Grimaldi, *NN approach and its comparison with NN-SVM to beta-barrel prediction*, Expert Systems with Applications, **61**, pp. 203–214 (2016).
7. I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas *et al*., *Machine Learning and Data Mining Methods in Diabetes Research*, Computational and Structural Biotechnology Journal, **15**, pp. 104–116 (2017).
8. F. Mansourypoor, S. Asadi, *Development of a Reinforcement Learning-based Evolutionary Fuzzy Rule-Based System for diabetes diagnosis*, Computers in Biology and Medicine, **91**, pp. 337–352 (2017).
9. A. Talaei-Khoei, J.M. Wilson, *Identifying people at risk of developing type 2 diabetes: A comparison of predictive analytics techniques and predictor variables*, International Journal of Medical Informatics, **119**, pp. 22–38 (2018).
10. N. Singh, P. Singh, D. Bhagat, *A rule extraction approach from support vector machines for diagnosing hypertension among diabetics*, Expert Systems with Applications, **130**, pp. 188–205 (2019).
11. M. Alirezaei, S.d Taghi, A. Niaki, S. Armin, A. Niaki, *A bi-objective hybrid optimization algorithm to reduce noise and data dimension in diabetes diagnosis using support vector machines*, Expert Systems with Applications, **127**, pp. 47–57 (2019).
12. U. Kanimozhi, S. Ganapathy, D. Manjula, A. Kannan, *An intelligent risk prediction system for breast cancer using fuzzy temporal rules*, National Academy Science and Letters,.**42**, *3*, pp. 227–232 (2019).
13. R.K. Bania, A. Halder, R-Ensembler, *A greedy rough set based ensemble attribute selection algorithm with kNN imputation for classification of medical data*, Computer Methods and Programs in Biomedicine, **184**, *105122* (2020).
14. N. Singh, P. Singh, *Stacking-based multi-objective evolutionary ensemble framework for prediction of diabetes mellitus*, Biocybernetics and Biomedical Engineering, **40**, *1*, pp. 1–22, (2020).
15. M.T. García-Ordás, C. Benavides, J.A. Benítez-Andrades, H.A. Moretón, I. García-Rodríguez, *Diabetes detection using deep learning techniques with oversampling and feature augmentation*, Computer Methods and Programs in Biomedicine, **202**, Article ID: 105968, (2021).
16. X. Xu, M. Yoneda, *Multitask air-quality prediction based on LSTM-autoencoder model*, IEEE Transactions on Cybernetics, **51**, 5, pp. 2577–2586 (2021).
17. P.M. Murphy, D.W. Aha, *UCI Repository of machine learning databases (machine-readable data repository*, University of California, Dept. Inf. Comput. Sci., Irvine, CA (1993).
18. S. Ganapathy, R. Sethukkarasi, P. Yogesh, P. Vijayakumar, A. Kannan, *An intelligent temporal pattern classification system using fuzzy temporal rules and particle swarm optimization*, Sadhana, **39**, *2*, pp. 283–302, (2014).
19. R.P. Cherian, N. Thomas, S. Venkitachalam, *Weight optimized neural network for heart disease prediction using hybrid lion plus particle swarm algorithm*, Journal of Biomedical Informatics, **110**, Article no. 103543, pp. 1–11 (2020).
20 R. Rajabioun, *Cuckoo optimization algorithm*, Applied Soft Computing, **11**, *8*, pp. 5508–5518 (2011).
21. H.T Gupta, P. Kumar, S. Saurabh, S.K. Mishra, B. Appasani, A. Pati, C. Ravariu, A. Srinivasulu, *Category boosting machine learning algorithm for breast cancer prediction*, Rev. Roum. Sci. Techn.–Électrotechn. et Énerg., **66**, *3*, pp. 201–206, Bucarest (2021).
22. M.I. Abdelwanis, R. El-Sehiemy, M.A. Hamida, *Parameter estimation of permanent magnet synchronous machines using particle swarm optimization algorithm*, Rev. Roum. Sci. Techn.–Électrotechn. et Énerg., **67**, *4*, pp. 377–382, Bucarest (2022).
23. V.E. Jesi, S.M. Aslam, *An intelligent disease prediction and monitoring system using feature selection, multi-neural network and fuzzy rules*, Neural Computing and Applications, **34**, *22*, pp. 19877–19893 (2022).