International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6520
***Article Received**: 18 February 2023 **Revised**: 20 March 2023 **Accepted**: 08 April 2023*

An Effective Disease Prediction System using CRF based Butterfly Optimization, Fuzzy Decision Tree and DBN

Manivannan D[1], Kavitha M[2]
[1]Department of Computer Science & Engineering,
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,
Chennai 600062, India
e-mail: mani02.ceg@gmail.com
[2]Department of Computer Science & Engineering,
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,
Chennai 600062, India
e-mail: mkavi277@gmail.com

Abstract— Diabetes is a seriously deadly disease today. It is necessary to enable patients to control their blood glucose levels. Even though, in the past, various researchers proposed numerous diabetic detection and prediction systems they are not fulfilling the requirements in terms of detection and prediction accuracy. Nowadays, diabetes patients are utilizing the gadgets like Wireless Insulin Pump that passes into the body instead of syringes for filling insulin. Within this context, insulin treatment is necessary for avoiding life-threatening. Toward this mission, a new deep learning approach-based disease detection system is introduced which takes care of identifying Type-1 and Type-2 diabetes, heart diseases, and breast cancer. In this system, a new Conditional Random Field based Butterfly Optimization Algorithm (CRF-BOA) is developedto select the important features for identifying the Type-1 and Type-2 diabetic disease. Besides, a new fuzzy ID3 classification method is developed for classifying the patient's datasets either normal or abnormal and disease affected. Ultimately, by applying the deep belief network (DBN) the classified patient records are involved with training to identify the relevant symptoms of similarity and glucose status of various patient records. These experiments are being conducted for proving the efficiency of the proposed deep learning approach in terms of glucose monitoring efficiency and disease prediction accuracy.The proposed approach achieved high detection accuracy than the current deep learning approaches in this directionbased on error rate and accuracy.

Keywords- Fuzzy ID3; CRF; BOA; DBN; Type-1 diabetic; Type-2 diabetic; Fuzzy Logic; Feature Selection; Classification.

I. INTRODUCTION

Diabetes is a widely spreading disease today in this world. The diabetes disease is classified as type-1 and type-2 diabetes that are based on a chronic condition where the human body doesn't produce sufficient insulin that results in fluctuations of blood glucose in human bodies. Because of the latest survey against this diabetic disease, around 41.5 crores of adults are diagnosed with type-1 and type-2 diabetes worldwide. Among them, 19.3 crores of adults are diabetes patients. The glucose level fluctuation of humans leads to serious health issues and life-threatening stages. Especially, type-2 diabetes affected patients must inject a particular amount of insulin daily to maintain their blood glucose level properly. This insulin treatment can detect only 50% to 60 % only. Hence, the blood glucose level monitoring process is the most important for all diabetic patients. Moreover, the number of diabetic patients increased to 42.2 crores. The main reason for getting affected by diabetes is the changes in people's food style and migration. Besides these, the blood glucose levels leave a high impact on the vision, kidney, mortality rate and weight of humans. By considering all these changes, the diabetes numbers may increase up to 75% in the year 2025 [2]. In this scenario, the inclusion of the blood glucose monitoring process is playing a major role in the diabetic disease prediction process.

Type-1 and Type 2 diabetes are serious deadly diseases today. Especially, Type-2 is a fourth major cause of mortality worldwide and also it became a major and common health issue that presents a significant occurrence. The difficulties are boosting the diabetes disease range due to an increase in the blood glucose level abnormally leads to damage to the molecules of a human body. Computer software is playing a major role in the process of medical diagnosis and also prognosis in the diagnostic process especially with the increasing huge volume of medical data. To overcome these disadvantages, a common structure-based model which adopts the fuzzy expert system under uncertainty for performing diabetic diseases with a deep learning approach is introduced in this work.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6520
***Article Received**: 18 February 2023 **Revised**: 20 March 2023 **Accepted**: 08 April 2023*

Feature selection is a preliminary work for performing the classification process in any prediction system and expert system. It is used to find the useful features from the input dataset. Generally, the feature selection process can be done under two major categories such as Filter approach and the wrapper approach. In both of these feature selection algorithms, we have incorporated the information gain ratio and the conditional probability for all the datasets. The filter method is used by various researchers who developed many emerging applications. Moreover, the information gain value, information gain ratio value and frequency of occurrences are considered by the researchers in their feature selection methods in the past to select the contributed features in the given input dataset. In this work, a CRF based BOA is developed to select the useful features effectively which combined two efficient feature selection algorithms.

Fuzzy logic is introduced and also adopted with various applications for deciding the various kinds of medical and network traffic datasets [5]. Fuzzy logic has been applied in the medical field for decision-making in 1985 [19]. Fuzzy set theory was applied in the next year 1986. After ten years, fuzzy logic got combined with a neural network for deciding the medical datasets. A fuzzy rule-based medical system was introduced [1] to predict the diseases which integrate the classification in data mining by applying fuzzy rules according to the fuzzy modelling. These kinds of fuzzy logic-based medical expert systems are useful for physicians to make the right decisions quickly over the medical records. Moreover, fuzzy inference systems were introduced and combined with neuro-fuzzy systems later on for diagnosing the diseases accurately in 2010. Besides, the fuzzy inference-based neuro-fuzzy system adopts the clustering methodology also for performing medical diagnoses.

Iterative Dichotomiser 3 (ID3) is a major classifier that was introduced in the year of 1983 and applied in emerging fields like economics and medical sciences in 1986. This ID3 classifier can handle a huge volume of data sets for performing classification. For performing classification in ID3, it uses the gain ratio and Gini-index for handling multi-value bias, partitioning the numerical attribute discretization. Here, the gain ratio is used for preferring the splits in unbalanced manner that is a partition and lower than others. The Gini-index is biased through multi-valued attributes. Moreover, it has disadvantages when handling a huge volume of data. However, identifying the optimal number of numeric features for performing partition is very difficult when presenting more numeric attributes in a dataset and it concentrates the time consumption for partitioning the numeric features.

Deep Learning is an emerging learning approach that can handle huge structured datasets and real-world datasets. The intelligence in the industry is heavily utilized for developing intelligent systems that apply neural networks including a back-propagation algorithm and are used for determining how a machine should alter its parameters to calculate the output correctly in every layer from the previous layer. This deep learning approach is used to handle the big data efficiently which has a volume of images, videos, speech files, and audio files [3]. This deep learning approach applies neural networks with different layers from input layer to output layer. Deep learning is adapted in various emerging applications for performing data-driven discovery processes. Many data processing mechanisms including Spark and Hadoop software and the standard statistical analysis tools like R are used. The deep learning approach is a new technique that is emerging to train and classify the data. The famous and successful deep learning-based emerging applications requires knowledge about the dataset, the algorithmic steps and efforts on programming, and significant computational resources. The major contribution of this paper is as below:

1. Propose a new deep learning technique for predicting diabetes as Type-1 and Type-2 by monitoring the glucose level continuously.
2. Introduce a new CRF based Butterfly Optimization Algorithm ICRF-BOA) to select the most important features for identifying the Type-1 and Type-2 diabetic disease, breast cancer, and heart diseases.
3. Propose a new fuzzy ID3 classification algorithm to categorize the patient records as “Normal” and “Disease affected”.
4. Finally, the classified patient records are involved with training for identifying the relevant symptoms similarities, and glucose status of various patient records by applying the deep belief network (DBN).
5. The proposed deep learning approach achieved high detection accuracy than the existing deep learning approaches.

This research paper is presented using six different sections. Among them, Section 1 is provided the introduction. Section 2 provides a brief survey about the different feature selection methods, classifications, fuzzy logic, deep learning, and deep belief network. The newly developed system is demonstrated in section 3. Section 4 provides the working flow and the newly proposed algorithms such as incremental feature selection, Fuzzy ID3, and Deep Belief Network. Section 5 shows the expertise of the proposed deep learning based on disease prediction accuracy. Section 6 is concluded the newly developed system and also suggests the possible works.

II. Related Works

The various disease prediction systems were developed by researchers to predict many diseases such as diabetes, cancer and heart [6,10,18]. Among them, a new genetic aware swarm

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6520
***Article Received**: 18 February 2023 **Revised**: 20 March 2023 **Accepted**: 08 April 2023*

method to extract the optimal features that are helpful for enhancing the classification accuracy [23,24,25]. They have evaluated their model with six different gene expressions by conducting experiments and achieved high classification accuracy. An extensive review about the various meta-heuristic methods [19,20]. The various methods are considered and analyzed the similarities and dissimilarities. Moreover, they have adopted the meta-heuristic methods with classifiers and enhance the prediction accuracy [22]. A fuzzy temporal cognitive map is developed and used [16] for predicting the various diseases such as Diabetes, Cancer and heart diseases. They have done a comparative analysis with other fuzzy modeling based on the accuracy obtained and proved as better. A new fuzzy min-max neural classifier with temporal constraints was constructed [17] and also incorporated a met heuristic technique called particle swarm optimization for optimizing the features that are capable of enhancing the classification result on heart, cancer and diabetes datasets. A new neuro-fuzzy classification algorithm was developed [21] for predicting the various diseases such as cancer, diabetes and heart. They have used the bell-shaped fuzzy membership method to perform the fuzzification process on input datasets by applying fuzzification matrix that input patterns were interrelated with classes. Finally, they have used neural classifiers with fuzzy logic and proved the effectiveness by obtaining superior accuracy to other models.

A new method to construct a rule based fuzzy classification based on the features of interpretability [4]. They have considered the fuzzy logics, fuzzy inference, fuzzy rules and the incorporation of various fuzzy membership functions to fix the fuzzy interval for generating the fuzzy rules to perform classification. Finally, they have done a comparative analysis between the various methods in terms of classification accuracy and misclassification rate. A detail analysis for determining the role of fuzzy logic over the decision-making process on various disease datasets [8]. A new DL technique which is capable of calculating the weights for the neural classifier which has 22 layers [1]. Their DL technique contains the various layers including dropout, rectified linear, convolution and pooling layers for enhancing the efficiency and effectiveness. Their DL is working opposite to the SVM and obtained better classification accuracy and provides better prediction result. A new multi-model stacked deep polynomial network that contains two different stages such as stacked deep polynomial network and multi-model data for performing feature fusion and feature learning to diagnose the Alzheimer disease and proved their system as superior to other systems by achieving better performance in classification [15].

A new fuzzy rule-based heart disease prediction system was developed and achieved better prediction result on standard heart datasets available in the UCI repository and proved as superior to other fuzzy rules-based classifiers. A new system to find the most significant features by using ML algorithms that are resulting to enhance the prediction accuracy in the process of detecting the heart disease and also achieved 88.7% as accuracy [9]. To identify the best disease prediction model from the existing ML and DL algorithms [13]. They have predicted the Type-1 diabetes diseases by using the selected ML and DL methods successfully and achieved better performance in terms of prediction accuracy. A new heart disease prediction system to detect the heart diseases [12]. Moreover, the proposed system consists of density aware spatial clustering method along with noise by detecting and eliminating the outliers. Moreover, they have used oversampling method for balancing the training data distribution process for predicting the heart diseases effectively by achieving 95.90% as prediction accuracy. A new heart disease diagnostic system has been developed by incorporating the various ML algorithms for making decisions on patient dataset [11]. They have incorporated the various ML algorithms including decision tree, neural classifier etc. In addition, their system uses the maximal relevancy and minimal redundancy to remove the less important features. The removal of less important features and the selection of more important features are useful for improving the classification accuracy and also reduce the execution time.

A new method is proposed for extracting the ECG related attributes that are useful for predicting the type of ECG that are collected from many patients as streaming data within a short span of time [14]. Their data is the collection of seven different ECG signals including Normal, Arrhythmia, etc. They have exploited the standard XG-Boost method to train the models that are obtained 99% as accuracy. A multi-model graph-based framework to predict the disease by incorporating the multi-model techniques that is capable of the rich data and also represent the learning process for aggregating the relevant features. A new ML based feature fusion technique was developed for predicting the diabetes diseases effectively by using SVM and ANN these are analyzed the patient detail and determines whether the patient affected with diabetes disease or not. They have used 70% of the dataset as training dataset and 30% of the dataset as testing dataset. Finally, their system obtained 94.87% as prediction accuracy which is greater than the accuracy obtained by other systems. To enhance the accuracy of their optimization method in the process of feature selection and the selected features are useful for achieving high classification accuracy of the neural classifier is 95.14%. Their model is obtained more than 1% higher accuracy than the standard neural classifier. A new heart disease prediction system by incorporating the optimized feature selection method. An IoT aware smart healthcare system to monitor the heart disease severity by using a meta-heuristic aware Fuzzy logic incorporated LSTM.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6520
***Article Received**: 18 February 2023 **Revised**: 20 March 2023 **Accepted**: 08 April 2023*

A new evolutionary method to select the necessary important features that are useful for predicting the diseases on high dimensional data were developed. A new uncertainty based temporal disease prediction system to detect the diseases dynamically also developed. A new regression tree aware classifier to predict the heart disease effectively was developed. A new feature selection approach by incorporating the mutual information and lower bound techniques for selecting the most important features that are capable of predicting the diseases. A relevance diversity method to select the contributed features those are helpful for enhancing the performance of their modified Bayes classification algorithm in terms of prediction accuracy. All the existing systems are not fulfilling the current requirements in terms of effectiveness and efficiency. For this purpose, a new prediction system is developed for predicting the diseases effectively with efficiency on diabetic, heart and cancer diseases.

III. System Architecure

The newly developed system architecture is demonstrating that the working procedure of the new system that contains various components including Dataset, User Interaction Agent, Feature selection Module, decision manager, rule manager and rule base. Here, the user interaction module collects the necessary patient records as input to perform pre-processing using feature selection module by decision manager. After selecting the features, the selected features are to be used for performing classification using classification module by the decision manager. The decisions are to be taken by using the rule manager suggestions and it uses the rules as well for finalizing the feature selection and classification processes.

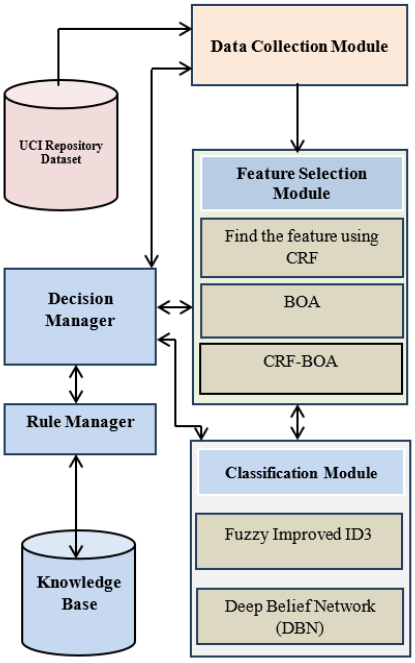

Figure 1. Proposed system architecture

The feature selection module applies CRF and BOA for selecting the useful features. Similarly, the classification module uses the Fuzzy Improved ID3 and DBN to categorize the patient records as "Normal" and "Diseased" by using the rules which are available in the knowledge based.

IV. Proposed Model

The newly developed CRF aware Butterfly Optimization Algorithm and the proposed Fuzzy aware Improved ID3 and DBN with necessary steps along with necessary background detail. This section describes the newly developed system.

A. Background

The background of the newly developed system is explained in this subsection. The background of Butterfly Optimization Algorithm (BOA) and Conditional Random Field (CRF) are provided. In addition, the fuzzy logic is also explained with necessary diagram and formulae. The BOA is explained first.

1) BOA

The BOA is a global optimization technique that is developed by Arora and Singh (2019). The BOA mimics the foraging the food and searching the mating partner attitudes of butterflies. The chemoreceptors of the butterfly are scattered over their bodies and also sense the receptors. Moreover, the butterflies are applied the chemoreceptors to sense or smile the flowers fragrance. In addition, the chemoreceptor is used for finding the optimal mating partner effectively. The butterfly is generating a fragrance when updates their locations. The movement of the butterflies in the BOA according to the fragrance intensity when it is failing for sensing the fragrance of any search space. The butterfly exploits the searching process by moving towards a position that is selected randomly. Here, the butterfly is used to sense the better movement towards the specific butterfly. In this process is called global search or exploration. The fragrance is also defined as a method. In BOA, the fragrance is well-defined as a stimulus intensity method and is demonstrated in equation (1) as below:

$$f = cIa \quad (1)$$

Where, the variable f indicates the butterfly fragrance magnitude, the variable c represents the sensory modality along with their values in between 0 and 1, I represent the fragrance intensity, and the variable a, indicates the power exponent which is based on the sensory modality along with their values between 0 and 1. Moreover, it is useful for controlling the fragrance degree by stimulus intensity. In addition, there are two major equations to update the positions of butterflies in BOA according to the value of f and t. The equation (1) is used to perform global search and it is also represented by equation (2) when perform the local search is shown in equation (3):

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6520
***Article Received**: 18 February 2023 **Revised**: 20 March 2023 **Accepted**: 08 April 2023*

$$x\ t{+}1\ i = x\ t\ i + r\ 2 \times g * - x\ t\ i \times fi, \quad (2)$$

$$x\ t{+}1\ i = x\ t\ i + r\ 2 \times x\ t\ j - x\ t\ k \times fi, \quad (3)$$

Where, the variable g∗indicates the best solution at present execution, the variable fi indicates the i^{th} butterfly fragrance, the variable r indicates a value between 0 and 1, the variable x t j indicates the j^{th} butterfly, and the variable x t k indicates the solution space of the k^{th} butterfly.

2) CRF

In CRF, the random variables are developed with the incorporation of conditional distributions by applying the models with conditions that have probabilistic properties. The CRF are useful to tag in sequence that are undirected graphic models with distributions P(b|a) directly to categorize the tasks. The CRF is a layered method in which each layer is considered as any one of the disease types. The probability value is computed for every attribute of the dataset. Moreover, the domain knowledge is also considered for the specific feature before selecting the layer that is independent with another layer. The Markov Model that is the base of CRF and it is also avoided the bias problems and not avoided by Markov model. Because of, the Moarkov model considers the conditional probability for a model with current state.

3) Fuzzy Logic

The basic concepts of fuzzy logic are described in detail in this section. Generally, it is explained the processing dealt using "degrees of truth" on behalf of "true or false", progressed by Zadeh in the 1960s. It gives a straightforward method for arriving at a clear conclusion based upon vague, uncertain, imprecise, noisy, or missing data. Fuzzy Logic uses a basic, rule-based IF x AND y THEN z method to resolve the control problems instead of endeavoring to demonstrate a framework numerically. Fuzzy expert system is a well-known computing structure according to the ideas, reasoning and IF-THEN rules. In general, the fuzzy inference comprises three major components namely the rule base, the physical database, and the inference engine. The rule base has some fuzzy rules which are used to perform the decision-making through inference. The physical database consists of facts that are necessary to make decisions by the inference engine by matching the IF part of the rules with the data present in the physical database. Moreover, the inference engine is used for making fuzzy decisions due to the uncertainty in the dataset by converting quantitative input into qualitative decisions. The inference engine performs deduction by applying inference rules using either a forward chaining inference mechanism or a backward chaining inference mechanism. The fuzzy inference system (FIS) is characterizing the membership functions which are utilized as a part of the fuzzy rules to perform the reasoning.

4) Fuzzification and defuzzification

The steps used by the Fuzzy Inference System are fuzzification, rule evaluation, aggregation of outputs from inference process using rules, and defuzzification. In fuzzification, the input values are given in the form of converted into qualitative values by applying the membership functions. The various functions such as Gaussian, triangular and trapezoidal membership functions are available to provide intervals that are used to fix the fuzzy rules for decision making. In the second step namely the rule evaluation step, the rules in the form of IF…THEN rules are evaluated by matching the IF part of the rules with the facts available in the knowledge base. For this purpose, a suitable inference mechanism is selected by the inference system. In some situations, a discriminant network is formed based on the queries given to the inference engine. The inference engine schedules the rules and executes them in an order specified by the scheduler. In the aggregation of outputs, the different outputs provided by the rules of a logical operator such as 'OR' are used to make the final decisions. In the decision-making process, the grouping is formed by choosing an aggregate function for finding either a maximum or minimum value from the group for comparison. In this process, new rules and facts may also be added to the knowledge base. In the final step called the de-fuzzification step, the fuzzy outputs are converted into crisp values by applying rules. Moreover, the values are converted to a normalized form to indicate the gradation of truth values. Hence, it helps to perform fuzzy and probabilistic reasoning. This work applies the Trapezoidal fuzzy membership function for making effective decisions.

B. Trapezoidal Fuzzy Membership Function

This section is explained the fuzzy logic and fuzzy rules with necessary detail. The fuzzy logic is helpful for generating the necessary fuzzy rules according to the deadline of the user requests and the size of the available resources. In this work, the standard trapezoidal fuzzy membership function is applied for generating rules. Generally, it contains four parameters such as a, b, c, and d which are demonstrated in equations (4) and (5).

$$trapezoid(x; a, b, c, d) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \frac{d-x}{d-c}, & c \le x \le d \\ 0, & d \le x \end{cases} \quad (4)$$

$$trapezoid(x; a, b, c, d) = \max\left(\min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0\right) \quad (5)$$

Here, the variables a, b, c and d along with a value that is less than b, b <= c, c < d that means which the x coordinate values of the four sides applied a membership function.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6520
***Article Received**: 18 February 2023 **Revised**: 20 March 2023 **Accepted**: 08 April 2023*

Moreover, it is useful for making decision on the patient records by applying the concern fuzzy rules. These rules are generated by following the fuzzy membership function intervals

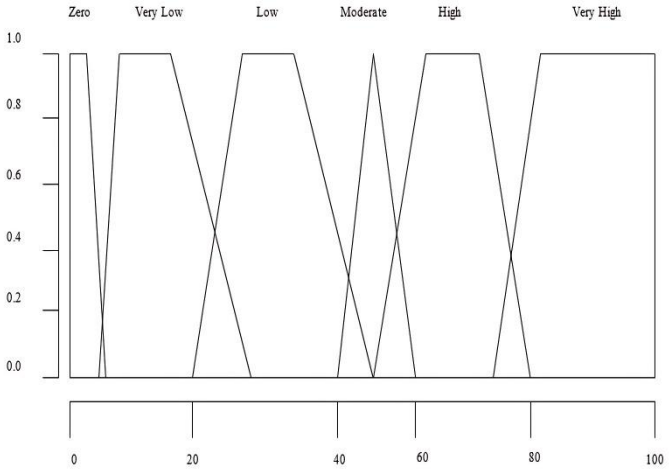

Figure 2. Fuzzy membership function

C. Feature Selection

The proposed CRF aware Butterfly Optimization Algorithm (CRF-BOA) is helpful for selecting the useful attributes from the standard medical datasets namely diabetic disease, breast cancer, and heart disease datasets. The various feature grouping techniques are working according to the differences between the features by using Intelligent CRF (Sannasi et al 2016). Moreover, the various steps of the CRF-BOA are as below:

Algorithm 1: CRF aware Butterfly Optimization Algorithm (CRF-BOA)

Step 1: Set butterflies positions Bi = (1, 2,3, …., N)
Step 2: Assign the values for the probability (PR), Modality_Sensory (CS), Power_Exponent (PE) and the maximum possible iterations (MI)
Step 3: While (! = MI)
Step 4: For the population of each BF
FV (F) = $SM.INT^{PE}$
// FV- Fragrance value, SM-Sensory Modality, INT-Intensity
Step 5: Find the best butterfly using ICRF (Sannasi et al 2016).
Step 6: For every population (BF) do
Step 7: Create a random number (rv) between 0 and 1.
Step 8: If (rv< p)
Perform exploration for bf position by using the equation
$x_i^{t+1} = x_i^t + (r^2 \times g^* - x_i^t) \times f_i$ // g^*- best solution
Else
Perform exploration by updatingthepositionof BF using
$x_i^{t+1} = x_i^t + (r^2 \times x_j^t - x_k^t) \times f_i$
// x_j^t and x_k^t are indicating the jth and kth BF.
End if
Step 9: Find the new_BF
Step 10: If the new_BF is betterthen
The population of the new BF must be updated.
Step 11: End for
Step 12: Update the PE value for the variable c.
Step 13: If the solution is better Then
Replace the best solution as result.
Step 14: End while
Step 15: Return the best solution
Step 16: Stop

In this incremental CRF aware BOA algorithm is helpful for identifying and selecting the useful features to perform effective classification. Moreover, apply the Intelligent CRF to select the useful features. The CRF is working based on the conditional probability. Here, it applies the fuzzy rules that are constructed by considering the CRF, IGR, and fuzzy intervals for finalizing the features. At the end, the more relevant features are grouped and also select the important features that can enhance the accuracy on benchmark dataset and hospital dataset.

D. Classification

This section discusses the classification process and it is also discussed in detail the proposed fuzzy ID3 classifier and the Deep Belief Network.

1) Fuzzy ID3

The Decision Tree (DT) is applied for making decision on input dataset through newly generated rules. These IF…THEN rules are generated by admin to make effective decision on different datasets. In this DT, the trapezoidal fuzzy membership is applied for generating rules to make decisions over the diabetic datasets. This work proposes a transform tree to fuzzy rules method for generating rule-based classification according to tree-based classifiers and also represents the fuzzy rules in the form of IF…. THEN.

Fuzzy logic is useful for making the exact decision on patient records in ID3. The newly developed Fuzzy ID3 is improved in terms of accuracy by applying effective fuzzy rules that are helpful to finalize the result with the disease severity level.

Algorithm 2: Fuzzy_ID3

Create_Fuzzy_Decision_tree (*TS*, List of features). *TS* is a training dataset.
Input: Training medical dataset
Output: Patient records with result
Step 1: Create a new node N
Step 2: IF the entire patient records in a specific group of patient records are in the similar group G THEN
Return "N" and mark it with class CL.
Step 3: the list of features is empty THEN
Return N as a leaf node and mark it with the majority class MCL
Step 4: IF the feature list is not empty THEN
Choose a feature with maximum IGR of feature as a test feature and mark node N as a test feature.
Step 5: the numeric feature is discretized according to the group feature values THEN, For each value *fi* of the test feature F
Step 6: Regard F= *fi* as a testing condition.
Step 7: Generates the relevant branch from the specific node N.
Step 8: Let *TSi*be the specific group when the value is present in F.
Step 9: IF*TSi= {}* THEN
Insert a new node as leaf and mark it with the maximum class in the input dataset Else
Store the return value of Create_Fuzzy_Decision_tree(*TSi*, {new list of features |{List of feature}−*F*}) in a list.
Step 10: Return rule-based decision list.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6520
***Article Received**: 18 February 2023 **Revised**: 20 March 2023 **Accepted**: 08 April 2023*

2) Deep Belief Network

The Deep Belief Network (DBN) is used to improve classification accuracy on diabetic datasets. Moreover, the DBN architecture and the necessary formulas that are applied in this DBN have been explained briefly with the necessary justification. It is useful for learning the probability distribution on medical datasets. In this paper, a DBN is developed and also applied RBM-based method to perform the classification effectively. In training process, the updated weights are used along with the gradient descent through the equations (6) and (7).

$$w_{ij}(t+1) = w_{ij}(t) + \eta \frac{\partial \log(p(v))}{\partial w_{ij}} \quad (6)$$

where(v) is the probability of a visible vector, which is given by:

$$p(v) = \frac{1}{Z}\sum_h e^{-E(v,h)} \quad (7)$$

The variable Z indicates the partition method that is used for performing normalization and $E(v; h)$ represents the energy function. The joint observation input is x and the hidden layer is modelled as below:

$$P(x, H_1, \ldots\ldots, H_N) = (\prod_{k=0}^{N-2} P(H_k|P_{k+1})).P(H_{N-1}, H_N) \quad (8)$$

Where, D $H0$, $P(H_k|H_{k+1})$ is a visible units with conditional distribution in the k number of layer hidden units conditions of Restricted Boltzmann Machine (RBM), the variable $P(H_{N-1}; H_N)$ indicates the visible-hidden joint distribution.

In the RBM, the input data is acquired from the first layer and it is also categorized as the data of the second layer. Moreover, two different ways can be selected for performing actions in an average manner. After performing the training process in an RBM and another one is stacked for adopting it and also considering these inputs from the trained layer which is a final layer. In addition, it initializes the new visible layer to perform a training vector and these are the values used to the various units of existing trained layers that are initialized by applying the necessary weights and biases. This entire process is to be repeated until reaches the endpoint. In the Backpropagation neural network (BPN), the last layer of the DBN is set as. In this scenario, the upper layers of RBM features are applied as an input vector that is for performing the training process over an entity classifier under supervision. Even though, each layer of RBM can ensure which its weight according to the feature vector is optimal after completion of the first level training process and the objective of the proposed model is for making the overall weight according to the optimized feature vector. Based on the categories of the standard BPN, it is capable of propagating error data from the upper layer to the lower layer of RBM and also achieved the adjusted DBN network. In this case, globalized optimization can be achieved. The various numbers of hidden layers and the possible number of neurons are present in each layer of DBN that are determined. The various steps of the algorithm are as follows:

Algorithm 3: DBN
Input: Test Data records
Output: Classified records
Step1: Assigned the population
Step 2: Generates the various hidden layers and the possible amount of neurons randomly in each layer.
Step 3: Computes the fitness value by using the equation (9).
Step 4: Choose the roulette function based on the fitness value.
Step 5: Perform the crossover and mutation operations alternately.
Step 6: Retains the ``Elite" and also retaining the greatest fitness valued individuals while conducting the evolution process.
Step7: Check whether it reaches the maximum number of iterations or not.
Step 8: If reached the maximum number of iterations then Retained the generated network structure Else Repeat steps 3-5.
Step 9: Apply the network which is optimized for the DBN and also train the prediction system.
Step 10: Categorize the testing dataset by using the trained DBN.
Step 11: Finally, compare the classification result with the classified test dataset for checking the classification accuracy.

V. EXPERIMENTAL RESULT ANALYSIS

This work has been developed by using Python programming language and it tested with WEKA tool by considering the standard benchmark medical datasets including heart, cancer and diabetes datasets that are taken from UCI Repository. In general, the disease prediction is important for physicians due to the arrival of patients and the introduction of new diseases.

A. Performance Metrics

The newly developed prediction system is evaluated by applying the metrics like precision value, recall value, f-measure value and prediction accuracy that are computed by applying the equations (9), (10), (11) and (12).

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (9)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (10)$$

$$F - Measure = \frac{(1+\beta^2) * Precision * Recall}{\beta * (Precision + Recall)} \quad (11)$$

$$\text{Prediction Accuracy} = \frac{\text{Number of records predicted successfully}}{\text{Total number of records available}} \quad (12)$$

B. Experimental Results

The effectiveness and efficiency of the newly developed prediction system is shown in this section by providing the experimental results that are conducted based on the various performance metrics. First, the newly developed CRF-BOA has

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6520
***Article Received**: 18 February 2023 **Revised**: 20 March 2023 **Accepted**: 08 April 2023*

been selected 6 important features as contributed features from heart disease from 15 features, 4 are identified as important features from diabetic dataset and 4 useful features are identified and selected from cancer dataset from 8 features. The proposed CRF-BOA takes very less time than the existing feature selection algorithms.

These important featured datasets are provided as input to the DBN for performing effective classification. The performance of the newly developed prediction system is evaluated by conducting various experiments and also compared with the standard classification algorithms such as C4.5, SVM, FTCM, TFMM-PSO and CNN. The effectiveness of the proposed model is proved by comparing with CNN. Figure 2 shows accuracy result over the different disease datasets like Diabetic, Cancer and Heart by considering the full dataset.

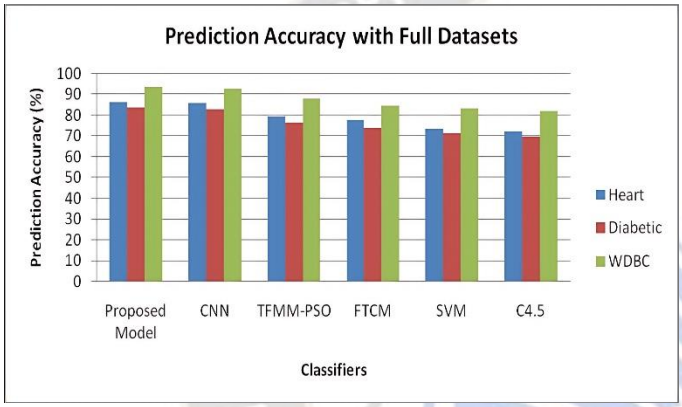

Figure 3. Accuracy with full datasets

From figure 3, the effectiveness of the classifiers of the newly developed disease prediction model that combines the Fuzzy ID3 and DBN is performed well than the standard classification algorithms such as C4.5, SVM, FTCM, TFMM-PSO and CNN. All the features of the disease dataset are used to carry out the experiments. The use of deep belief network and fuzzy rules are the reasons for the performance enhancement in this work.

Figure 4 demonstrates the accuracy over the reduced heart, diabetic and cancer datasets by considering the selected features of the datasets.

From figure 4, it demonstrates the effectiveness of the newly developed disease prediction accuracy is proved that by comparing the performance with the available classification algorithms namely CNN, TFMM-PSO, FTCM, SVM and C4.5 according to the disease prediction accuracy over the reduced featured datasets. The newly developed is achieved better accuracy than other systems even by considering the selected features contained datasets.

Table 1 demonstrates that the comparative study according to the performance of the newly proposed disease prediction model on full features contained datasets and the selected features contained datasets in terms of prediction accuracy.

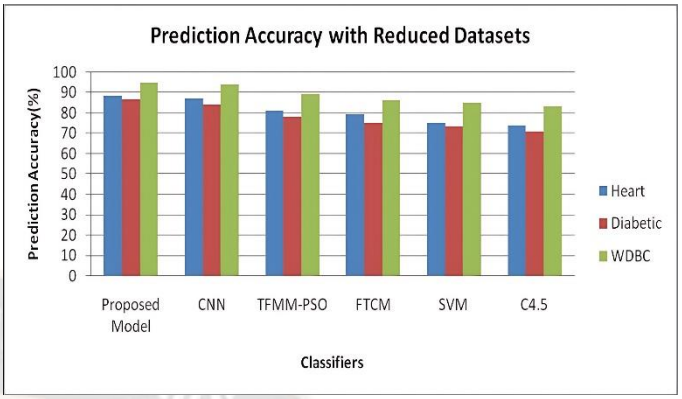

Figure 4. Accuracy on the reduced disease datasets

TABLE I. COMPARATIVE ANALYSIS

Disease Datasets	**Accuracy (%)**	
	Full features Dataset	***Selected Features Contained Dataset***
Heart Disease Dataset	86.1	87.5
Diabetic Disease Dataset	83.1	84.6
WDBC Disease Dataset	92.9	94.4

TABLE II. TIME ANALYSIS

Disease Datasets	**Proposed Model (CRF-BOA + Fuzzy ID3 + DBN)**		**Fuzzy Temporal Cognitive Map**	
	Training Time (sec)	***Testing Time (sec)***	***Training Time (sec)***	***Testing Time (sec)***
Heart Disease Dataset	0.38	0.19	0.42	0.20
Diabetic Disease Dataset	1.69	0.79	1.77	0.84
WDBC Disease Dataset	0.40	0.11	0.43	0.14

The prediction accuracy considered comparative analysis is done in the work and demonstrated in table 1. Here, the three datasets like diabetic, cancer and heart are used as full featured dataset and selected featured dataset.

The time taken analysis is demonstrated in table 2 that considers the 3 datasets like heart, diabetic and cancer disease datasets. Here, the time analysis is done by performing the training and the testing the different inputs.

From table 2, it is proved that the newly developed system's efficiency according to the training and testing time on three datasets than the available classifiers including Fuzzy Temporal

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6520
***Article Received**: 18 February 2023 **Revised**: 20 March 2023 **Accepted**: 08 April 2023*

Cognitive Map that incorporates the fuzzy temporal rules on decision making process.

Figure 5 demonstrates the accuracy analysis by considering the newly developed system and the domain experts. The domain experts are able to predict the diabetic disease, heart disease and cancer diseases. This analysis considers the16 different number of patient records such as 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500 and 1600.

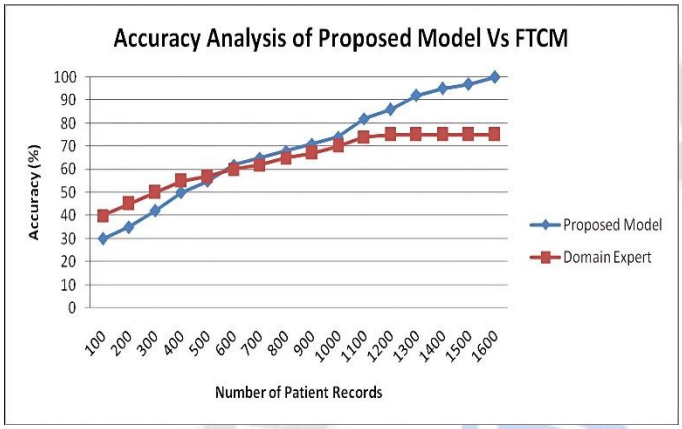

Figure 5. Accuracy analysis w.r.t proposed system and domain expert

From figure 5, it is observed that the newly developed system's prediction accuracy is superior to domain expert's prediction result when considers the similar group of inputs to conduct the experiments. Almost the prediction accuracy on various datasets is equal from tenth experiment onwards. The last 7 experiments results are down and stabilized for the two different. The reason for the enhancement is the application of optimization technique and fuzzy logic.

The proposed system is tested by considering the standard metrics in this performance analysis which is demonstrated in table 3. This analysis is done by using the diabetic dataset, heart dataset and cancer dataset.

TABLE III. PERFORMANCE ANALYSIS

Disease Datasets	Performance Metrics		
	Precision Value (%)	*Recall Value (%)*	*F-Measure Value (%)*
Heart	87.01	98.43	93.96
Diabetic	85.21	99.37	94.65
Cancer	94.90	98.71	94.30

Table 3 is proved the newly developed system's effectiveness is evaluated in this result by considering the metrics with the use of DBN, feature selection technique and fuzzy ID3. The reason for the performance enhancement is the application of fuzzy logic, deep learning approach and the effective feature optimization technique that uses Butterfly Optimization and CRF.

The prediction accuracy analysis is done by considering the performance of the newly developed system and the available neural networks-based classifiers with fuzzy logic and temporal constraints that are developed. In this analysis 10 experiments have been done with various sizes of records start from 100 to 1000 records.

Figure 6 proved the efficiency of the newly developed system by obtaining high accuracy which is superior to the available neural classifiers with the incorporation of fuzzy temporal logic. The reason for the enhancement is the application of CRF and Butterfly Optimization processes, fuzzy rules along with decision tree and the deep belief network.

Figure 7 shows the error rate (misclassification rate) for the newly developed system and the available neural classifiers with the incorporation of fuzzy logic and temporal logic. In this analysis, five different experiments have been done on various set of patient records.

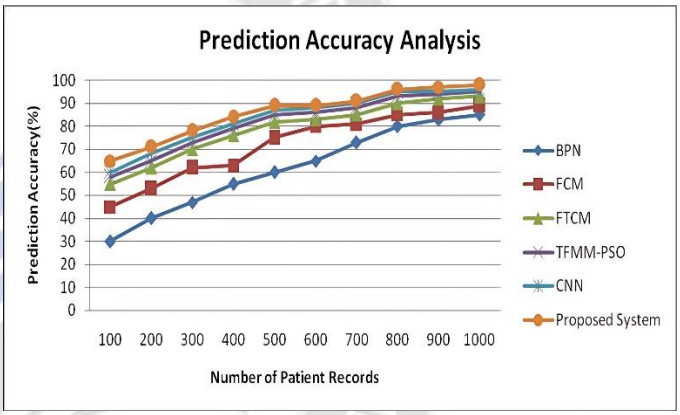

Figure 6. Prediction accuracy analysis

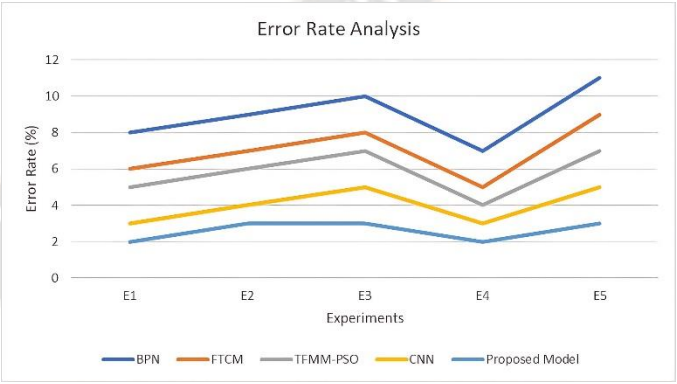

Figure 7. Error Rate Analysis

From figure 7, it is seen that the newly developed system's error rate is superior to the available neural classifiers with the incorporation of fuzzy logic and temporal logic. The reason for the enhancement is the application of CRF, BOA, fuzzy logic, Euclidean distance that is incorporated in decision tree and DBN.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6520
***Article Received**: 18 February 2023 **Revised**: 20 March 2023 **Accepted**: 08 April 2023*

VI. Conclusions And Future Direction

The proposed system has been developed to predict the diseases such as Type-1 and Type-2 diabetes, heart diseases, and breast cancer. In this system, a new Conditional Random Field based Butterfly Optimization Algorithm (CRF-BOA) is incorporated to select the important features for identifying the Type-1 and Type-2 diabetic disease. Moreover, a new fuzzy ID3 classification method is also developed to classify the patient's records as “Normal” or “Diseased”. Ultimately, the classified patient records are involved with the training process to identify the relevant symptoms of similarity and glucose status of various patient records by applying DBN. The experiments have been done to prove the efficiency of the newly developed deep classifier in terms of glucose monitoring efficiency and disease prediction accuracy**.** Finally, the proposed system is achieved high detection accuracy than the current deep learning techniques in this direction based on error rate and accuracy. This work can be enhanced with the introduction of new optimized deep classifier for reducing the training and testing time.

References

[1] MainakBiswas, VenkatanareshbabuKuppili, Damodar Reddy Edla, Harman S. Suri, Luca Saba, RuiTatoMarinhoe, J. Miguel Sanches, Jasjit S. Suri, "Symtosis: A liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm", Computer Methods and Programs in Biomedicine, Vol.155, pp. 165-177, 2018.

[2] NoureenTalpur, Said JadidAbdulkadir, HithamAlhussian, MohdHilmiHasan, MohdHafizulAfifi Abdullah, "Optimizing deep neuro-fuzzy classifier with a novel evolutionary arithmetic optimization algorithm", Journal of Computational Science, Vol.64, No.101867, 2022.

[3] Marco Pota, Massimo Esposito, Giuseppe De Pietro, "Designing rule-based fuzzy systems for classification in medicine", Knowledge-Based Systems, Vol.124, pp. 105-132, 2017.

[4] IlhemBoussaïd, JulienLepagnot, Patrick Siarry, "A survey on optimization metaheuristics", Information Sciences, Vol.237, pp. 82-117, 2013.

[5] HosseinAhmadi, MarsaGholamzadeh, Leila Shahmoradi, MehrbakhshNilashi, PooriaRashvand, "Diseases diagnosis using fuzzy logic methods: A systematic and meta-analysis review", Computer Methods and Programs in Biomedicine, Vol.161, pp. 145-172, 2018.

[6] P. Ganesh Kumar, T. Aruldoss Albert Victoire, P. Renukadevi, D. Devaraj, "Design of fuzzy expert system for microarray data classification using a novel Genetic Swarm Algorithm", Expert Systems with Applications, Vol.39, No.2, pp. 1811-1821, 2012.

[7] SoumadipGhosh, SushantaBiswas, DebasreeSarkar, ParthaPratimSarkar, "A novel Neuro-fuzzy classification technique for data mining", Egyptian Informatics Journal, Vol.15, No.3, pp. 129-147, 2014.

[8] U. Ahmed et al., "Prediction of Diabetes Empowered With Fused Machine Learning", IEEE Access, vol. 10, pp. 8529-8538, 2022.

[9] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", IEEE Access, vol. 7, pp. 81542-81554, 2019.

[10] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare", IEEE Access, vol. 8, pp. 107562-107582, 2020.

[11] S. Zheng et al., "Multi-Modal Graph Learning for Disease Prediction", IEEE Transactions on Medical Imaging, vol. 41, no. 9, pp. 2207-2216, Sept. 2022.

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

[13] J. Xie and Q. Wang, "Benchmarking Machine Learning Algorithms on Blood Glucose Prediction for Type I Diabetes in Comparison with Classical Time-Series Models", IEEE Transactions on Biomedical Engineering, vol. 67, no. 11, pp. 3101-3124, Nov. 2020.

[14] D. Bertsimas, L. Mingardi and B. Stellato, "Machine Learning for Real-Time Heart Disease Prediction", IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 9, pp. 3627-3637, Sept. 2021.

[15] Perez-Siguas, R. ., Matta-Solis, H. ., Matta-Solis, E. ., Matta-Perez, H. ., Cruzata-Martinez, A. . and Meneses-Claudio, B. . (2023) “Management of an Automatic System to Generate Reports on the Attendance Control of Teachers in a Educational Center”, International Journal on Recent and Innovation Trends in Computing and Communication, 11(2), pp. 20–26. doi: 10.17762/ijritcc.v11i2.6106.

[16] 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, Springer, Vol. 39, No.2, pp. 283-302, 2014.

[17] 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, IOS Press, Vol. 26, No.3, pp. 1167-1178, 2014.

[18] U.Kanimozhi, S.Ganapathy, D.Manjula, A.Kannan, "An Intelligent Risk Prediction System for Breast Cancer Using Fuzzy Temporal Rules", National Academic Science Letters, Vol.42, No.3, pp. 227-232, 2019.

[19] Meneses-Claudio, B. ., Perez-Siguas, R. ., Matta-Solis, H. ., Matta-Solis, E. ., Matta-Perez, H. ., Cruzata-Martinez, A. ., Saberbein-Muñoz, J. . and Salinas-Cruz, M. . (2023) “Automatic System for Detecting Pathologies in the Respiratory System for the Care of Patients with Bronchial Asthma Visualized by Computerized Radiography”, International Journal on Recent and Innovation Trends in Computing and Communication, 11(2), pp. 27–34. doi: 10.17762/ijritcc.v11i2.6107.

[20] Jothi CS, Ravikumar S, Kumar A, Suresh A. An approach for verifying correctness of web service compositions. International Journal of Engineering &Technology. 2018;7(1.7):5-10.

[21] Ravikumar, S. and Kannan, E., 2022. Machine Learning Techniques for Identifying Fetal Risk During Pregnancy. International Journal of Image and Graphics, 22(05), p.2250045.

[22] Ravikumar, S. and Kannan, E., 2022. Analysis on Mental Stress of Professionals and Pregnant Women Using Machine Learning

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v11i5.6520
***Article Received**: 18 February 2023 **Revised**: 20 March 2023 **Accepted**: 08 April 2023*

Techniques. International Journal of Image and Graphics, p.2350038.

[23] MJ, C.M.B., Arif, M. and V, D.K., 2022. Linguistic Analysis of Hindi-English Mixed Tweets for Depression Detection. Journal of Mathematics, 2022, pp.1-7.

[24] Kannan, E., Ravikumar, S., Anitha, A., Kumar, S.A. and Vijayasarathy, M., 2021. Analyzing uncertainty in cardiotocogram data for the prediction of fetal risks based on machine learning techniques using rough set. Journal of Ambient Intelligence and Humanized Computing, pp.1-13.

[25] Antony Kumar, K. and Carmel Mary Belinda, M.J., 2022. A Multi-Layer Acoustic Neural Network-Based Intelligent Early Diagnosis System for Rheumatic Heart Disease. International Journal of Image and Graphics, (p.2450012).