Identification of Female Infertility in People with Thalassemia using Neural Network

Mohammad Ordouei*  
Computer Engineering Dep, Islamic Azad University, South Tehran Branch, Tehran, Iran  
dr.ordouei@gmail.com  
ORCID: 0000-0002-0391-0638

Abstract — Female infertility has become one of the basic problems in families in the last few decades. Infertility is one of the most important social problems that has a structural impact on modern society. Despite the significant advances in thalassemia pathobiology and the efficacy of chelation regimens, complications of transfusion therapy have attenuated the reproductive health of thalassemia patients. The aim of this article is to obtain a model of the integration of a neural network and genetic algorithm that can identify and predict its effective parameters before the onset of infertility in people with thalassemia to prevent this disease. Another goal of the article is to identify infertility in people with thalassemia with relatively high accuracy and speed using the parameters given to the neural network. The classification accuracy for the perceptron neural network is 96.8254% for test data.

Keywords — Data mining, Genetic algorithm, Neural network, Thalassemia Introduction

I. INTRODUCTION

Intelligent systems, as systems that can help humans in various fields, have been highly regarded in the last few decades. Artificial intelligence, multi-agent systems, data mining, and machine learning are among the fields of interest in all sciences [1]. Data mining means extracting information and knowledge and discovering hidden patterns from a very large database, and these patterns and knowledge are usually hidden in the data. Data mining can be used to perform tasks such as clustering, categorizing, and predicting data [2]. It also has many applications in various scientific fields such as banking, disease diagnosis, pattern recognition, and penetration detection [3].

Thalassemia is a genetic disorder in which hemoglobin is lost its normal structure and the body produces non-useful hemoglobin. Thalassemia includes two types alpha and beta. The alpha type of this disease is silent, which is also called beta thalassemia minor. Beta thalassemia is one of the most common quantitative disorders of hemoglobin in the world, especially in Iran [4]. Hypogonadotropic hypogonadism (HH) is the most common endocrine disorder in women with thalassemia major who receive blood transfusions. 51 to 66% of thalassemia patients with marked hemosiderosis are susceptible to puberty failure, sexual dysfunction, infertility, and short stature. People with thalassemia may also suffer from iron accumulation in the ovaries or testes, and oxidative stress may occur there when there is an imbalance between the production of reactive oxygen species and the capacity to scavenge antioxidants in the reproductive tract [5,6].

Infertility is often higher in women with thalassemia who need blood transfusions. However, some women with the disease are able to become pregnant naturally. Of course, it is possible to be pregnant with thalassemia. Women with thalassemia should often use drugs to stimulate ovulation to become pregnant. Many of the physical problems caused by thalassemia are related to the accumulation of large amounts of iron in the body. This extra iron can accumulate in the body due to the disease itself or frequent blood transfusions needed to treat thalassemia. The accumulation of iron causes many problems in different organs of the body and affects parts of the body such as the pituitary gland and the hypothalamus responsible for the production of hormones, thereby preventing ovulation [7].

Infertility in thalassemia patients is caused by iron deposition in the pituitary gland, which disturbs the pituitary-gonadal axis in 40-90% of patients with regular transfusions. Patients with this condition present with pubertal failure symptoms, such as breast, uterus, and vaginal atrophy [8]. Ovulation induction should only be undertaken by a specialist reproductive team, according to Human Fertilization and Embryology Authority (HFEA) guidelines, keeping in mind and making women aware of the risk of hyperstimulation syndrome, multiple pregnancy, ectopic pregnancy, and miscarriage. Around 1%–2% of induced ovulation cases develop severe hyperstimulation syndrome causing fluid retention with bloating, breathlessness, and nausea, resulting in abdominal pain, vomiting, dyspnea, and rapid weight gain, while the most severe cases have to be hospitalized because of hypovolemic shock, renal and/or respiratory insufficiency, and arterial thromboembolism [9].

In the continuation of this research, in order to get more familiar with the proposed concepts, in the second part, the background of the study is discussed, and in the third part, the proposed research method or methodology is stated. In the fourth part, the data analysis of the proposed method is mentioned.

II. RESEARCH BACKGROUND

Known neural networks are used to produce very accurate results in medicine it is also a powerful tool to help doctors analyze, model, and generate complex clinical data in a wide range of applications. These networks are a set of input and output connections where each connection has its own weight. One of the most famous types of artificial neural networks is the multi-layer perceptron1 neural network.

These networks include several layers of nodes called perceptrons, which are connected to each other by a fully connected feed-forward network. Neurons of all layers except the input layer (which is linear) are modeled as neurons with nonlinear Activation function2 [10].

Article History:  
Received Date: Jan. 05, 2023, Accepted Date: Mar. 21, 2023, Available Online: Apr. 01, 2023

5371
Roby et al stated that heart disease may be one of the leading causes of death. Due to the lack of knowledge and experience of experts about the symptoms of heart failure, it is not easy to diagnose the disease. As a result, computer-based prediction of heart disease patients can play an important role in pre-stage diagnosis to take appropriate measures regarding the recovery of patients.

However, choosing the appropriate classification method of data mining can effectively predict the initial stage of the disease to return from it.

In this paper, we review three commonly used classification techniques such as support vector machine (SVM), k nearest neighbor (KNN), and artificial neural network (ANN), according to their evaluation for prediction. The incidence of heart disease has been studied using the standard kidney disease data set.

- Multi-Layer Perceptron
- Activation Function

Experimental results show that the classification accuracy using SVM is 85.1852\% (%) better than using KNN (82.663\%) and ANN (73.333\%) [11].

Hansen et al. stated that the knowledge about the interactions of drugs with each other usually "comes from premature tests and contradictory drug reports or based on the knowledge of the mechanism of action".

The aim was to check if drug interactions with each other can be discovered without prior hypotheses using data mining.

Six drug groups were discovered with known effects that increased the INR, and two had a known effect in a highly related drug class (oripavine [buprenorphine] derivatives and natural opium alkaloids).

Anti-propulsive had an unknown signal of INR increase [12]. Using artificial intelligence methods, Yousufian and colleagues tried to identify effective parameters in thalassemia disease in their collected data.

They used neural network and decision tree methods and reached 95.48 and 96.65 accuracy [13].

The structure of the proposed method in this article is shown in the form of a flowchart (Figure 1).

The criteria designed for this research are based on an Excel file, and the information related to the time of each of the obtained criteria is collected based on the documentation in the clinic files from 2017 to 2018. The criteria and options related to infertility and thalassemia disease in women according to their values are shown in Table 1.

<table>
<thead>
<tr>
<th>ROW</th>
<th>Standard title</th>
<th>Standard title</th>
<th>Standard title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>First name</td>
<td>Type of iron removal treatment</td>
<td>PTH</td>
</tr>
<tr>
<td>2</td>
<td>Last name</td>
<td>Use of hydroxyurea</td>
<td>VinB</td>
</tr>
<tr>
<td>3</td>
<td>Age</td>
<td>Liver LEC</td>
<td>Extradial</td>
</tr>
<tr>
<td>4</td>
<td>Gender</td>
<td>Card MRI</td>
<td>Perispot</td>
</tr>
<tr>
<td>5</td>
<td>Type of thalassemia</td>
<td>C-Care</td>
<td>Terfenac</td>
</tr>
<tr>
<td>6</td>
<td>Age of marriage</td>
<td>EkoEF</td>
<td>HBA</td>
</tr>
<tr>
<td>7</td>
<td>Spouse's health status</td>
<td>EkoPAP</td>
<td>HeF</td>
</tr>
<tr>
<td>8</td>
<td>Spontaneous pregnancy</td>
<td>Liver MRI</td>
<td>HBA</td>
</tr>
<tr>
<td>9</td>
<td>Pregnancy with medication</td>
<td>InSM</td>
<td>HBB</td>
</tr>
<tr>
<td>10</td>
<td>Abortion</td>
<td>Femoral density</td>
<td>HBA2</td>
</tr>
<tr>
<td>11</td>
<td>Number of IUT</td>
<td>Spinal bone density</td>
<td>HCVab</td>
</tr>
<tr>
<td>12</td>
<td>Number of children</td>
<td>WBC</td>
<td>Hapridic</td>
</tr>
<tr>
<td>13</td>
<td>Sets practices</td>
<td>Hbs</td>
<td>Antibody (Albick)</td>
</tr>
<tr>
<td>14</td>
<td>Alpha genetics</td>
<td>Platelet</td>
<td>Antibody (atenu)</td>
</tr>
<tr>
<td>15</td>
<td>X-mal</td>
<td>FBS</td>
<td>Ovary sone</td>
</tr>
<tr>
<td>16</td>
<td>Blood group</td>
<td>Uric acid</td>
<td>Bone uterus</td>
</tr>
<tr>
<td>17</td>
<td>Blood RR</td>
<td>REN</td>
<td>Date of birth</td>
</tr>
<tr>
<td>18</td>
<td>Height</td>
<td>Cr</td>
<td>The proceeding</td>
</tr>
<tr>
<td>19</td>
<td>Weight</td>
<td>SGOT</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Blood transfusion age</td>
<td>SGOT</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Transfusion</td>
<td>Ferritin</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Blood transfusion interval</td>
<td>Ca</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Splenectomy</td>
<td>LH</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Age of splenectomy</td>
<td>FSH</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Maturity state</td>
<td>YSH</td>
<td></td>
</tr>
</tbody>
</table>

The criteria and options for the connection between infertility and thalassemia are shown in Table 1.
III. The proposed method

In the combination of perceptron neural network and genetic algorithm, the purpose of neural network training by a genetic algorithm is to find the best values of weights for neurons so that the neural network error is minimized for them.

Therefore, training should be done optimally. In the genetic algorithm, each answer (population member) has a vector consisting of weights.

The objective function is used to evaluate each vector. In the initial preparation stage, each member of the population includes a vector containing the weights of the neural network, which is sent as an input to the objective function after it is randomly assigned.

The objective function receives the input vector including the weights and considers it as the weights of the neural network. Then the training inputs are applied to the constructed network and based on the output of the neural network and the target output of the network, the error is calculated and returned.

In the repetition stage, each member of the population, which includes the weight, is changed by the mutation and intersection operators in the algorithm and is evaluated again by the objective function.

The repetition continues until the termination conditions are met. After completing the training process, the algorithm returns the member of the population that has the best values of weights. The member selected by the algorithm is the member whose mean square error of the neural network is lower than the rest of the population.

A. Evaluation of research criteria based on the perceptron neural network method

In Table 1, variables 1 to 54 are predictor variables and variable 55 is the target variable on which the prediction results are based and it is divided into two values determined at the time of thalassemia patients who are fertile or not. The data were analyzed in the three-layer perceptron neural network based on 70% training data and 30% experimental data. The results of these criteria are shown in Table 2.

The number of hidden layers of the perceptron neural network is one layer.

The learning rate is 0.3.

The number of learning steps is equal to 500 and the validation threshold is 20.

We use the Back propagation1 algorithm for the perceptron neural network. In this method, the adjustment coefficients (synaptic weights) are updated every time after a complete review of the data of a training course. In the proposed method, we use the combination of a genetic algorithm and a three-layer perceptron neural network. The number of data is 209, 70% of which were used as training data and 30% as test data. The population size in the genetic algorithm is set to 100. The maximum number of generations in the genetic algorithm is 100. The percentage of the intersection operator is 0.5. The percentage of mutation operators is 0.35.

B. Simulation comparison of the proposed models

Due to the use of C4.5 decision tree algorithms, a three-layer perceptron neural network, and the proposed method of combining genetic algorithm and a three-layer perceptron neural network, there was confidence in the evaluation of the model. The prediction accuracy in this research with 70% training data and 30% test data, for the C4.5 decision tree method, is equal to 99.978% and for the perceptron neural network is equal to 96.8254% and for a genetic algorithm with perceptron neural network, it was equal to 99.9724%.

The results of the comparison of the number of correct samples are shown in Figure 2.

In examining the research criteria, the obtained Information Gain value shows that among the 63 criteria, 20 criteria have priority in the forecasting process. Spontaneous pregnancy is the first criterion and is very effective in predicting thalassemia disease, and the number of children is the second most effective criterion in predicting thalassemia disease. The priority of the research criteria is shown in Figure 3.

IV. Conclusion

The results show that the root mean square error in the perceptron neural network method is equal to 0.0248 and for the combined genetic algorithm and three-layer perceptron neural network is 0.0276. The classification accuracy for the C4.5 decision tree method, perceptron neural network, and combined genetic algorithm and three-layer perceptron neural network is 99.978%, 96.8254%, and 99.9724% respectively for the test data.
REFERENCES


