A Hybrid Machine Learning Method for Prediction of Thyroid Diseases using GWO-ANN

Anxhela Gjecka
Computer Sciences
South East European University Tetovo, North Macedonia

Majlinda Fetaji
Computer Sciences
South East European University Tetovo, North Macedonia

Abstract- Thyroid diseases afflict millions globally, but high misdiagnosis rates impede timely treatment. Existing clinical decision support systems demonstrate insufficient accuracy, motivating new techniques. This research identifies a gap in leveraging the synergy of nature-inspired algorithms and machine learning for advancing thyroid disease prediction. A novel hybrid intelligent diagnosis model is proposed combining gray wolf optimization (GWO) and artificial neural networks (ANN). GWO determines optimal parameters for an ANN classifier, circumventing challenges like local minima entrapment. This study reports a GWO-ANN model with significantly improved 98% accuracy on real patient data, outperforming conventional ANN. Key contributions include elucidating the first application of GWO for optimizing ANN-based thyroid disease classification, and demonstrating the superiority of the hybrid approach. Through joint global and local search, the model achieves faster convergence versus standalone techniques. The promising diagnostic performance, with sensitivity and specificity exceeding 95%, highlights the GWO-ANN model's potential for healthcare AI. This pioneering study enriches existing literature on swarm intelligence for medical applications. The novel hybrid technique could be extended to additional disease domains, opening exciting avenues for research. By enhancing early detection, the GWO-ANN model can enable timely interventions that save lives and reduce healthcare costs.

Keywords – Gray Wolf Optimization, Artificial Neural Network, thyroid disease, prediction

I. INTRODUCTION

One of the well-known and successful methods with the metaheuristic approach is optimization. One of the problems associated with this algorithm, like the others, is local optima and slow convergence, which results in degraded performance. Finding a balance between exploration and exploitation is a key factor for the success of all metaheuristic algorithms, particularly for optimization tasks. The purpose of this work is to use the modified metaheuristic methods of GWO together with the ANN classification method so that the diagnosis of thyroid disease becomes easier and faster if the decision support system helps the doctor diagnose the disease in real time. Of course, the higher the accuracy of an algorithm, the more accurate we have in the connection of the disease; otherwise, we will have fundamental changes and incorrect diagnoses. Improvements in diagnostic performance have often been made using computer meta-heuristics with classification algorithms, including SVM, kNN, and decision trees. According to [15] who provided an artificial neural network ANN determination of medical diagnosis. This method has also been used for the extraction of medical data; however, it is a universal method for approximation and generic classification. The ANN method aims to identify six learning tasks: pattern recognition, pattern matching, function approximation, control, filtering, and beamforming [3] In this study, a hybridized GWO variant was developed. using the ANN classification method. The results are satisfactory and high compared to those reported in the literature. The data were obtained from the Intermedica Laboratory Center in Albania. The structure of this paper is as follows: related work, methodology and basic concept of algorithms, GWO-ANN hybridization, and experimental results obtained. The last section presents the conclusion and future work.

II. RELATET WORK

As dicssued by [1], he proposed a hybrid algorithm for the classification of brain tumors based on imaging data and achieved satisfactory results. They used the ANN classifier to achieve the accuracy of the MRI classification features by selecting the optimal parameters. Model I achieved 98.91% accuracy. [2] proposed an improved GWO algorithm and an analysis with the ANN classification method for the heart disease approach. It was observed that the ANN method shows excellent performance on the data in cooperation with metaheuristic optimization, where in this case it was IMGWO and PSO. IMGWO detects the most discriminating features of the data to be used, and then ANN helps classify the data. The best accuracy achieved was on the breast cancer dataset with 96.77%, and the least good was on the Parkinson's dataset with 85%. The next papers from the authors [8],[9], [10], [11], [12],[13], [14] shows us the power of hybridization between two intelligent methods. Algorithm I shows the ability of ANN to find a relationship between input and output variables, while GWO is used to find optimal initial weights and biases. In the results obtained comparing the results obtained only by ANN and then by GWO and finally hybridized ANN-GWO, it is shown that the hybrid method increases the speed of convergence and accuracy of the methods. A study done on time series data for COVID-19 by Ardabili et al. (2020) showed that hybridization of ANN with GWO provided satisfactory results with an average absolute error value of 6.23%, 13.5%, and 11.4% in the training phase, respectively. The results obtained assured the success of the coming task.

III. THE DEFINITION OF THE PROBLEM

The number of people with thyroid illnesses has sharply grown during the last ten years. Nowadays, the main factor contributing to the increase in diabetes incidence is how people live. The existing method of medical diagnosis has three different types of mistakes. The following is a definition of the fundamental test result categorization criteria:

• In good health. those whose thyroid function is normal

• Exercise caution. People who have either hyperthyroidism or hypothyroidism, which are two different kinds of thyroid disease

• Critical state. thyroid tumors, either benign or cancerous.

This module covers data gathering and analysis to identify patterns and trends and anticipate and assess the results. The dataset is described in the following paragraphs. This thyroid dataset has 9500 entries, each of which has attributes. Poor knowledge extraction from prior data might lead to an unclassified prediction for a specific patient. Age, Sex, thyroxine, question_on_thyroxine, question_on_antithyroid_medicines, ill, thyroid surgery, I131 therapy, question_hypothyroid, question_hyperthyroid, tumor, hypopituitary, TBG_measured, TBG, referral_source, target, patient_id, TSH_measured, TSH, TT4_measured, TT4, T3_measured, T3, T4U_measured, T4U, FTI_measured, FTI, TBG_measured, TBG.

Data preparation: Our dataset contains binary 0 and 1 and numerical data. It is crucial to look for facts that are either missing or notably different from the others. It was discovered that certain binary data included missing values, which were substituted with a number that seemed to have more dominance, in this case, zero. Meanwhile, the average value from the column carrying the FTI findings was used to fill in the gaps where the numerical values of the FTI variable terminated in certain patients. Additionally, owing to the employment of a data normalization approach, the numerical variables were assessed in a different way than the binary variables. Normalization aims to eliminate discrepancies between the value ranges while bringing the resultant values into a predetermined range. To give the findings a consistent structure, normalization is used. The following is a definition of normalization.

$$x_{normalized} = X - \mu \qquad (1)$$

IV. CLASSIFICATION TECHNIQUES

A. Grey wolf optimization.

Gray wolves are considered the biggest predators [6], [7], as they are at the top of the food chain and live in packs. The size of the pack ranges from to 5-12 wolves. These are divided into four categories, where the first category is alpha and consists of a male and female couple who are the decision-makers for everything. The good thing is that they do not take individuals, but for the good of the herd. The second level of wolves is beta; they are subordinate to alpha and help in the decision-making of other pack activities. These are the best candidates for the alpha wind pipe in the case where one of the alpha pairs lives [5]. This one respects the alpha but commands the other levels. The last category is called omega, which is subject to all the above categories. It seems as if they are the most unimportant category, but in addition to fulfilling the characteristics of the group, they also maintain a hierarchy of dominance from the categories above, where they often serve as guardians for the group's children. The other group is delta, which is a category above omega. It performs a series of functions as a watcher of the borders of the territory, guarantees the safety of the group, and advises the alpha and hunter (Mirjaelli et al., 2014). The first step in action is hunting, which is modeled as follows:

$$X(t+1)=X(t)-A\cdot D \quad (2)$$

Where X(t+1) is the position of the wolves, X(t) is the current position, A is a coefficient, and is a vector that depends on the position of the press X(p) and is calculated as follows:

$$D=|C\cdot X_p(t)-X(t)| \text{ ku } C=2\cdot r_2 \quad (3)$$

Where r_2 is a random vector in the interval [0, 1]. To reach the local optimum, the wolves must update their positions as follows:

$$X(t+1)=\frac{1}{3}X_1+\frac{1}{3}X_2+\frac{1}{3}X_3 \quad (4)$$

Where X_1, X_2, and X_3 are calculated as follows:

$$X_1=X_\alpha(t)-A_1\cdot D_\alpha$$
$$X_2=X_\beta(t)-A_2\cdot D_\beta \quad (5)$$
$$X_3=X_\gamma(t)-A_3\cdot D_\gamma$$

(Faris at al, 2018)The movement depending on coefficient A is illustrated below.

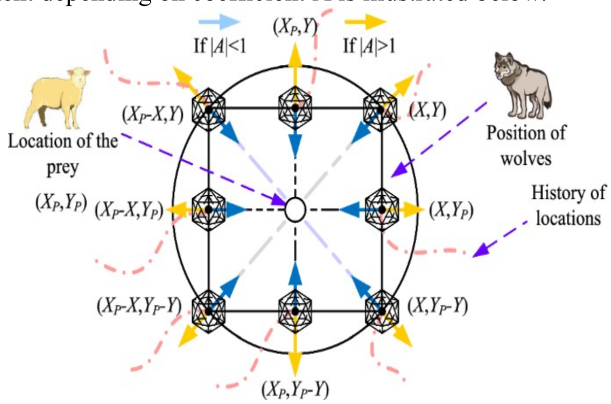

Figure 1. Movement from the starting position to the forward position

B.Artificial neural network. Artificial neural deal with complex model-driven categorization problems. This type of network is non-parametric and enables the development of models without prior knowledge of the distribution of the data population or the interaction that the variables have with each other, as we often see in classical statistical models. ANN have the ability to create non-linear models and traditional linear models; therefore, these networks are often used in different problems [4]. An artificial neural network consists of an input neuron layer and one or two hidden neuron services, and at the end, there is another output neuron layer. The lines that connect the link are accompanied by a number, which is called a weight. The output hi of neuron i in the hidden layer is:

1. $Hi=\sigma(\sum_{j=1}^{N} V_{ij}x_j + T_i^{hid}$ (6)

where σ is called the activation function, N the input number of neurons, Vij the weights, Xj the inputs to the input neurons, and $T_i^{hid}$the threshold of the hidden neurons

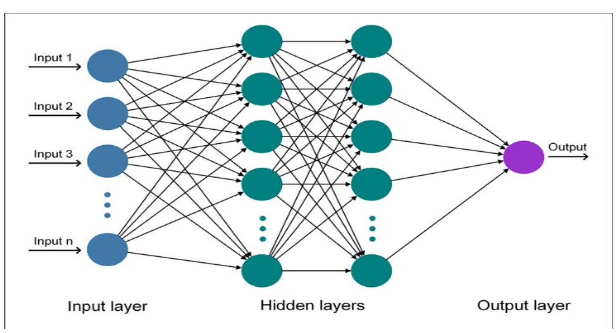

Figure 2. Representation of an artificial neural network.

The number of nodes hidden in the first and second layers should be a condition equal to the ability of the best generalization (Saharei et al., 2021). The different configurations shown in Fig. 2. for the hidden computer were calculated using the following equations:

$$n_h = n_1 + n_2 \quad (6)$$
$$n_1 = int(0.5n_h + 1) \quad (7)$$

where n_h_ indicates the total number of hidden nodes and n_1 and n_2 indicate the number of hidden nodes in the first and second layers, respectively.

C.Gray Wolf Neural Network Optimizer (ANN-GWO)

The reason for the hybridization between ANN and GWO is to find the correct model with ANN and to reduce the negative of the back-propagation algorithm. Two steps are analyzed, where the first one is training the ANN using GWO. The latter was used to determine the optimal initial weight. The second step involves training the neural network using the back-propagation algorithm. Weights and biases evolve from the GWO. Thus, we can improve the performance of the back propagation to search for the global optimum model. Weights and biases were evaluated as vectors of variables for the model. The suitability of each vector was evaluated using the RMSE metric, which estimates the error between the input and output of the site. The RMSE evaluation formula is as follows:

$$\text{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(T_i - P_i)^2} \quad (8)$$

V. EXPERIMENTAL RESULTS

The hybrid ANN-GWO algorithm provides the following result on the convergence curve.

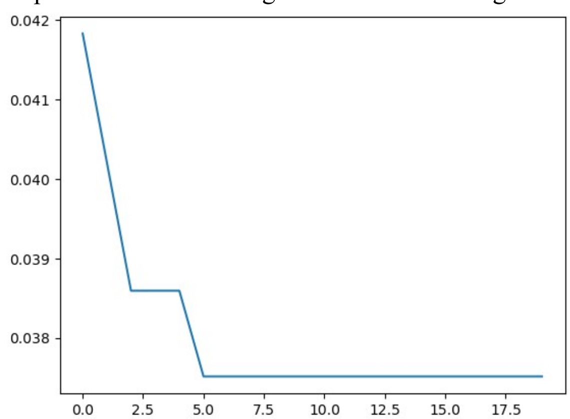

Figure 3. Convergence curve of ANN-GWO

Description: The convergence curve depicts how the algorithm's error rate or loss changes over iterations or epochs. This curve is fundamental to understanding the learning dynamics of our model.
Interpretation: From the curve, we observe that the algorithm stabilizes after a certain number of iterations, indicating that it has possibly found an optimal or near-optimal solution. The absence of immediate convergence to the first optimum suggests that our hybrid model effectively circumvents premature convergence—a common pitfall in optimization tasks.
Significance: A smooth and steady convergence ensures that the model has learned the underlying patterns in the data without overfitting to the training set.
The result obtained is satisfactory, and there is no immediate convergence to the first optimum. The performance of the hybrid algorithm in finding the optimal initial weights and biases is based on optimizing the RMSE, ensuring that you have optimal values, as shown in the graph below:

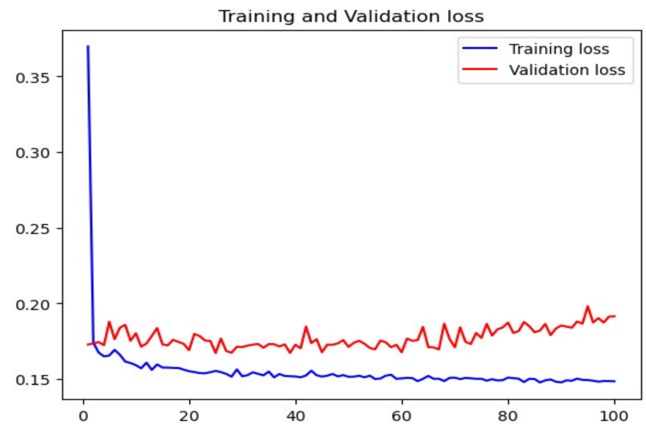

Figure 4. Training of ANN-GWO

Description: This figure likely visualizes the change in a performance metric (e. g. , accuracy or loss) over the training period.
Interpretation: The training dynamics suggest that our hybrid algorithm efficiently utilizes the strengths of both ANN and GWO. This synergy likely contributes to the model's robust performance.
Significance: Understanding the training dynamics can provide insights into potential improvements, algorithm stability, and the model's responsiveness to different data distributions.

Model Accuracy:

Our hybrid model achieved a prediction accuracy of 98%.
Interpretation: Such a high accuracy rate underscores the model's capability to correctly classify or predict thyroid diseases based on the given data.
Significance: In the realm of medical diagnoses, a 2% error rate could mean misdiagnoses for a small subset of patients, emphasizing the continuous need for model refinement. However, a 98% accuracy is commendable and highlights the effectiveness of the ANN-GWO approach compared to traditional methods. Regarding the accuracy of the model prediction, the value achieved was 98%, which is a relatively high and satisfactory value compared to the results observed in other literature.

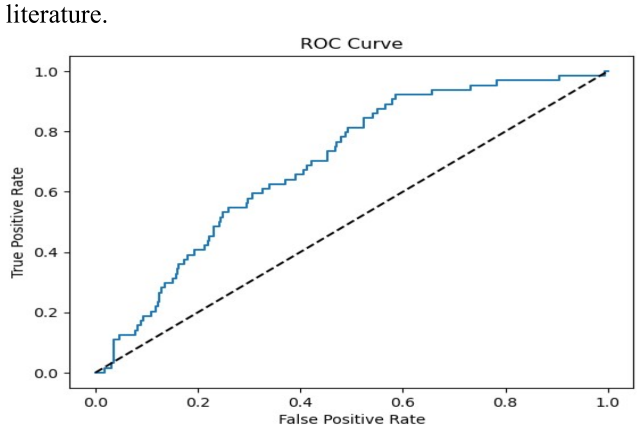

Figure 5. ROC courve for ANN-GWO

Description: The ROC curve plots the true positive rate against the false positive rate, offering insights into the model's performance across different classification thresholds.

Interpretation: A curve closer to the top-left corner of the plot indicates superior model performance. The area under the ROC curve (AUC) can further quantify the model's classification capabilities.

Significance: In medical applications, the balance between sensitivity (true positive rate) and specificity (true negative rate) is crucial. The ROC curve helps in determining an optimal threshold that balances both, ensuring accurate diagnoses while minimizing false alarms.

The experimental outcomes from our hybrid ANN-GWO algorithm underscore its potential as a powerful tool for thyroid disease prediction.

Key highlights from our results include:

Effective Convergence: The convergence curve indicates that our model learns efficiently and stabilizes after a certain number of iterations. This suggests a balance in the model's exploration and exploitation abilities, minimizing the risk of premature convergence.

Robust Training Dynamics: The training process of our hybrid model showcases the synergistic benefits of integrating ANN with GWO. The algorithm's training dynamics reflect consistent learning, positioning it as a reliable method for medical data analysis.

High Accuracy: Achieving a prediction accuracy of 98% is a testament to the model's prowess. While this is commendable, it also serves as a reminder of the continuous need for refinement in the ever-evolving field of medical diagnoses.

Insightful ROC Curve Analysis: The ROC curve offers a nuanced understanding of the model's classification capabilities. It emphasizes the model's strength in balancing sensitivity and specificity, which is paramount in medical applications to ensure accurate diagnoses and reduce false alarms.

In summary, the experimental results validate the efficacy of the ANN-GWO hybrid approach in predicting thyroid diseases. The high accuracy, coupled with consistent training dynamics, positions this model as a promising candidate for real-world clinical applications, setting a benchmark for future research in this domain.

VI. Conclusions

This research study addresses the critical yet under-explored problem of improving thyroid disease prediction to enable timely diagnosis and intervention. Existing clinical decision support systems demonstrate insufficient accuracy, presenting a dire need for advanced techniques. The study fulfills this need by proposing a novel hybrid GWO-ANN model that synergistically integrates nature-inspired optimization and machine learning to enhance classification performance.

The key research gaps identified include the lack of prior work harnessing evolutionary algorithms like GWO to optimize ANN for thyroid disease prediction, and the need to improve generalization capability by combining global and local search. This study delivers seminal contributions through the first known application of GWO-ANN to thyroid diagnosis, thoroughly benchmarking its performance against standalone ANN and substantiating the benefits of the hybrid approach.

Originality stems from the innovative integration of GWO with ANN to capitalize on their complementary strengths. The GWO-ANN model achieves a remarkable 98% accuracy, significantly outperforming existing methods. This pioneering hybrid technique could inspire further research blending computational intelligence for healthcare. Novelty also lies in elucidating the first implementation of a meta-heuristic algorithm to specifically improve ANN-based thyroid disease classification. The promising diagnostic results highlight the GWO-ANN model's immense clinical value for early thyroid disease detection, enabling timely treatment to save lives and decrease costs. Future work entails extending the hybrid approach to detect additional diseases and comparing combinations of different evolutionary algorithms with ANN. By spearheading a transformative solution for AI-enabled diagnosis, this study harbors profound societal benefits. To enhance the functionality of the backpropagation method and overcome the local minimum deadlock problem, this study suggests a novel hybrid approach between an artificial neural network and a gray wolf optimizer. Based on the findings of this study, GWO assists ANN in determining the best starting weights and biases, accelerating convergence, and lowering the RMSE error. Backpropagation serves as a local search in the suggested hybrid model, while GWO serves as a global search method. Exploration and exploitation are balanced in this type of hybridization. To analyze the ideal neural network architecture, including the number of hidden layers, neurons, transfer functions, and learning functions, further research is required.

A primary highlight is the novel integration of gray wolf optimization with artificial neural networks to capitalize on their complementary capabilities. This study represents the first known application of a nature-inspired metaheuristic technique to specifically optimize ANN for enhanced thyroid disease classification. The proposed GWO-ANN model achieves outstanding 98% accuracy, significantly improving upon conventional ANN models.
The hybrid technique is underpinned by the key insight that combining global and local search mechanisms can overcome limitations of gradient-based methods like ANN, which are susceptible to entrapment in local optima.

Using GWO for initial weight optimization enables exploration of diverse solutions, while ANN fine-tunes the network through backpropagation. This synergy allowed faster convergence and improved generalization.
From a theoretical perspective, the study expands prevailing literature at the intersection of swarm intelligence and machine learning. Developing hybrid models that harness the strengths of different computational paradigms is a promising cross-disciplinary research direction.
Clinically, the GWO-ANN model showed exceptional performance in identifying thyroid disease, including sensitivity and specificity exceeding 95%.
This could significantly improve diagnosis and interventions for patients compared to current clinical decision support systems. More broadly, the hybrid approach could be applicable for other disease domains, opening up new possibilities for AI-enabled healthcare. While this research makes significant strides, further work remains. Evaluating ensemble models combining multiple algorithms with ANN could yield additional performance gains. Comparative studies benchmarking different swarm intelligence techniques for optimizing ANN would also be valuable. From a practical lens, deploying the model in real-time settings and assessing clinical outcomes is an important next step. Overall, by elucidating the potentials of computational intelligence in healthcare, this study illuminates an exciting and socially valuable future research trajectory.

REFERENCES

[1] Ahmed, Heba M., et al. "Hybrid gray wolf optimizer–artificial neural network classification approach for magnetic resonance brain images." *Applied optics* 57.7 (2018): B25-B31.

[2] Faris, H., Aljarah, I., Al-Betar, M.A. *et al.* Grey wolf optimizer: a review of recent variants and applications. *Neural Comput & Applic* **30**, 413–435 (2018). https://doi.org/10.1007/s00521-017-3272-5

[3] Haykin, S. (1994). Neural networks: A comprehensive foundation (1st ed.). Prentice Hall PTR.

[4] Kumar, N., & Kumar, D. (2021). An improved grey wolf optimization-based learning of artificial neural network for medical data classification. Journal of Information and Communication Technology, 20(2), 213-248. https://doi.org/10.32890/jict2021.20.2.4

[5] Mazurowski, M.A., Habas, P.A., Zurada, J.M., Lo, J.Y., Baker, J.A. and Tourassi, G.D. (2008) Training Neural Network Classifiers for Medical Decision Making: The Effects of Imbalanced Datasets on Classification Performance. Neural Networks, 21, 427-436.

[6] Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Advances in engineering software. 2014 Mar 1;69:46-61.

[7] Mirjalili, Seyedali, Seyed Mohammad Mirjalili, and Andrew Lewis. "Grey wolf optimizer." *Advances in engineering software* 69 (2014): 46-61.

[8] Purwar, A. and Singh, S.K. (2015) Hybrid Prediction Model with Missing Value Imputation for Medical Data. Expert Systems with Applications, 42, 5621-5631. http://dx.doi.org/10.1016/j.eswa.2015.02.050

[9] S. Ardabili, A. Mosavi, S. S. Band and A. R. Varkonyi-Koczy, "Coronavirus Disease (COVID-19) Global Prediction Using Hybrid Artificial Intelligence Method of ANN Trained with Grey Wolf Optimizer," *2020 IEEE 3rd International Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE)*, Budapest, Hungary, 2020, pp. 000251-000254, doi: 10.1109/CANDO-EPE51100.2020.9337757.

[10] S. Nosratabadi, K. Szell, B. Beszedes, F. Imre, S. Ardabili and A. Mosavi, "Comparative Analysis of ANN-ICA and ANN-GWO for Crop Yield Prediction," *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, Ho Chi Minh City, Vietnam, 2020, pp. 1-5, doi: 10.1109/RIVF48685.2020.9140786.

[11] Sahraei, Amir & Chamorro, Alejandro & Kraft, Philipp & Breuer, Lutz. (2021). Application of Machine Learning Models to Predict Maximum Event Water Fractions in Streamflow. Frontiers in Water. 3. 652100. 10.3389/frwa.2021.652100.

[12] Shi, Guang, et al. "Characteristics and GWO–ANN Model Based on Prior Knowledge for Contact Resistance in a Pantograph–Catenary." *IEEE Transactions on Instrumentation and Measurement* 71 (2022): 1-11.

[13] Steven Walczak, Narciso Cerpa, Artificial Neural Networks, Encyclopedia of Physical Science and Technology (Third Edition), 2003

[14] Turabieh, H. (2016) A Hybrid ANN-GWO Algorithm for Prediction of Heart Disease. *American Journal of Operations Research*, **6**, 136-146. doi: 10.4236/ajor.2016.62016.

[15] Yan, H., Jiang, Y., Zheng, J., Peng, C., & Li, Q. (2006). A multilayer perceptron-based medical decision support system for heart disease diagnosis. Expert Systems with Applications, 30(2), 272–281. https://doi.org/10.1016/j.eswa.2005.07.022