

# HIGHLY ACCURATE HYBRID METHOD FOR ATTENTION DEFICIT HYPERACTIVITY DISORDER CLASSIFICATION BASED ON ANFIS-RFE-GWO

Deepika\*, Shaveta Arora and Meghna Sharma

The NorthCap University, Department of Computer Science and Engineering  
Gurugram, Haryana, India

DOI: 10.7906/indecs.23.6.3  
Regular article

Received: 10 July 2024.  
Accepted: 26 November 2025.

## ABSTRACT

Attention Deficit Hyperactivity Disorder (ADHD) is one of the most prevalent neurobehavioral disorders, characterised by persistent patterns of inattention, impulsivity, and restlessness. This disorder significantly affects the personal, social, and academic development of individuals, with millions of children and adolescents worldwide experiencing its symptoms. Despite its high prevalence, accurate diagnosis is still a significant challenge for medical professionals, as distinguishing affected individuals from healthy controls is often complex. Many machine learning and deep learning approaches have been proposed for its diagnosis; the accuracy of ADHD diagnosis is still insufficient and needs further improvement. This study proposes a highly accurate hybrid framework to address this gap. The proposed method integrates the Recursive Feature Elimination technique to select the relevant and most significant feature subset, an Adaptive Neuro-Fuzzy Inference System to perform classification while handling inherent data uncertainty, and Grey Wolf Optimisation for hyperparameter tuning. Cross-validation is further applied to ensure optimal feature subset selection and robust model performance. The proposed framework is evaluated using phenotypic data from the ADHD-200 dataset, which included 547 patients diagnosed with the disorder and 325 healthy controls. For performance benchmarking, the proposed model is compared with several conventional machine learning classifiers, including Random Forest Classifier, K-Nearest Neighbour, and Gradient Boosting Classifier etc. Experimental results demonstrate that the proposed model outperforms several state-of-the-art approaches, achieving a classification accuracy of 98,30%, sensitivity of 96,82%, and F1-score of 97,60%. These results highlight the model's potential as a reliable and effective diagnostic tool for clinical decision support, performing accurate detection and reducing misdiagnosis.

## KEY WORDS

attention deficit hyperactivity disorder, adaptive neuro-fuzzy inference system, fuzzy logic, grey wolf optimisation, recursive feature elimination

## CLASSIFICATION

JEL: I12

PACS: 87.57.R-

\*Corresponding author,  $\eta$ : [deepikagarg1310@yahoo.co.in](mailto:deepikagarg1310@yahoo.co.in); +91 9582895328;  
157, Sec-14, Gurugram, 122001, Haryana, India

## **INTRODUCTION**

Attention Deficit Hyperactivity Disorder (ADHD) is a prominent neuro-behavioural ailment of the modern age. It affects 5-9% of kids and adolescents worldwide [1] and may persist in adulthood as well. ADHD is the third most common mental disorder after depression and anxiety. Various studies indicate a significant increase in ADHD global prevalence from 1997 to 2022. An alarmingly high prevalence rate of 9,4% was observed in US kids in 2016 [2], and approximately 366,33 million adult ADHD cases were reported worldwide in 2020 [3]. The prevalence of ADHD among young children further increased to 10,47% in 2022 [4]. ADHD has frequent comorbidity with several other mental disorders, like schizophrenia, anxiety, depression, and sleep disorders [5-7]. Lack of concentration, increased activity level, and impulsiveness are the most noticeable symptoms of ADHD. Based on the symptoms, three different categories of ADHD have been identified, namely: i) Predominantly inattentive, ii) Predominantly hyperactive, and iii) ADHD-combined. ADHD imposes significant adverse effects on individuals. Children affected by ADHD are more likely to engage in disruptive conduct at school, face problems in making friends and relationships, and fulfil official commitments. They also have poor academic performance, resulting in a higher dropout rate in graduation and higher studies. Such children easily get engaged in increased usage of drugs, suicide attempts, and several other life-impairing activities, leading to a high risk of increased mortality rate. ADHD places a substantial financial burden of an estimated \$122,8 billion on society [8] due to unemployment, productivity loss, and healthcare service requirements.

## **RESEARCH MOTIVATION**

At present, there is no individual lab test or objective diagnostic method to diagnose ADHD. The diagnosis process mainly relies on the behavioural symptoms [9] observed by parents and teachers. Many questionnaire-based rating scales are used to assess the IQ level of the patient. The collected data is then analysed by medical experts to interpret the results. This approach is quite subjective as there is no golden standard for the assessment of results. Medical experts render diagnoses based on their experience, knowledge, and personal interpretation. Different medical experts may provide different diagnoses, often leading to misdiagnoses/overdiagnoses of ADHD. Another significant challenge in diagnosing ADHD arises due to the uncertainty linked to its symptoms. ADHD is heterogeneous; different patients can exhibit different symptoms altogether, posing complexities for accurate diagnosis by healthcare professionals. The presence of overlapping symptoms further compounds the diagnostic uncertainty. Various ADHD symptoms overlap with several other psychological disorders (e.g., schizophrenia, anxiety, depression, etc.), blurring the lines between symptoms and complicating the diagnostic process. Hence, there is a pressing need for an automated diagnostic process that is objective, precise, and reliable for ADHD, given its global prevalence and severity of impairments. Many studies [10-12] have been proposed for ADHD classification based on patients' clinical data, yet the desired level of accuracy remains to be achieved. This is primarily due to the discrete approach adopted in most of the existing studies for ADHD classification. These studies have not factored in the inherent uncertainty in the data. Owing to heterogeneity and overlapping of its symptoms, it is not possible to quantify ADHD data accurately, resulting in substantial uncertainty and ambiguity within the dataset. This study seeks to effectively tackle these uncertainties and put forward a highly precise and efficient model for ADHD classification.

## **MAJOR CONTRIBUTIONS**

This study proposes a highly accurate, reliable, and efficient hybrid model based on metaheuristic and fuzzy logic for ADHD classification. It utilises the personal characteristic

data present in the benchmark ADHD-200 dataset [13]. Medical datasets generally contain many redundant or insignificant features that could create noise in the data and substantially degrade the classification accuracy. Feature selection plays an important role in removing such features. The proposed model employs an efficient feature selection technique, Recursive Feature Elimination (RFE) [14], to eliminate the irrelevant features and generate an optimal feature set, yielding the highest performance.

Uncertainty associated with ADHD data can be addressed using fuzzy logic, which is an excellent method for handling ambiguity and inconsistencies. It effectively addresses inaccurate boundaries by exploring all the intermediate possibilities between absolute true and absolute false in the form of IF-THEN rules. Fuzzy logic has been widely used in a variety of industries, including education, medicine [15], decision support systems, image processing [16], supply chains and forecasts [17], etc. However, identifying the correct set of rules for fuzzy logic is very important, which requires a domain expert with a vast understanding of the system. To overcome this limitation, the present study combines fuzzy logic with neural networks. Neural networks have good pattern-matching and learning capabilities, so these can be used for the development of fuzzy rules, eliminating the requirement for expert knowledge. This combination is known as the Adaptive Neuro-Fuzzy Inference System (ANFIS) [18]. ANFIS has demonstrated superior performance compared to traditional fuzzy logic-based systems for accurately tackling a wide range of difficult real-world issues. By integrating the adaptability and learning capabilities of neural networks along with the interpretability and human-like reasoning of fuzzy logic, ANFIS has gained a lot of popularity among research communities. Furthermore, ANFIS enables efficient optimisation and parameter tuning through its trainable parameters, resulting in improved accuracy and robustness. Numerous efficient classification systems have been developed using ANFIS for the detection of medical disorders like depression [19], breast tumour identification, Alzheimer's detection [20], corona detection [21], and heart disease prediction, etc.

Like all other neural networks, ANFIS also exhibits certain limitations. Random initialisation of weights, biases, and other network parameters can sometimes lead to erratic model behaviour, where the model's performance fluctuates irregularly during the training process. Such instability can make it difficult to converge to an optimal solution, consequently influencing the classification performance of the network. So, fine-tuning the network parameters is crucial for optimising the network performance. Combining ANFIS with an optimisation method will yield the most favourable network parameters and improve classification accuracy. Considering the same, this study uses a reliable metaheuristic technique named Grey Wolf Optimisation (GWO) [22] to optimise various hyperparameters of ANFIS. GWO is a population-based optimisation technique that continuously explores the hyperparameter space, evaluates different combinations of hyperparameters against some objective function, and returns the best possible combination of hyperparameters. There are many other popular optimisation algorithms, such as Particle Swarm Optimisation (PSO), Genetic Algorithm (GA), Artificial Bee Colony (ABC), etc. GWO exhibits considerably better robustness than these algorithms [23]. Additionally, it offers faster convergence with less computational burden [24]. Another reason for selecting GWO is its strong ability to balance exploration and exploitation in high-dimensional, non-linear search spaces. Since the proposed framework involves optimising multiple ANFIS parameters, GWO's efficiency in avoiding local minima while maintaining computational efficiency [25, 26] makes it particularly well-suited for this problem. Hence, the combination of ANFIS-GWO has the potential to create a resilient classification system demonstrating superior classification accuracy. The proposed system is also validated on a diverse set of disease datasets to establish its efficacy. The remarkable performance of the model on these datasets highlights its adaptability and robustness in handling the complexity inherent in various types of medical data.

The key contribution of this article can be summarised as follows:

- A highly accurate, reliable, and efficient hybrid model based on fuzzy logic and metaheuristics is proposed for ADHD classification.
- The model is implemented using the patient's clinical data only, which is in accordance with the standard diagnostic process of ADHD.
- The issue of missing feature data is effectively handled by using a reliable imputation technique called Multiple Imputation by Chained Equation (MICE) [27, 28].
- An optimal feature subset comprising the most significant/discriminative features is selected using a robust RFE technique with cross-validation.
- A fuzzy logic-based classifier, ANFIS, is used to manage the uncertainty associated with data.
- The model hyperparameters are fine-tuned using an efficient optimisation technique, GWO, resulting in a remarkable enhancement in model performance.
- The proposed model achieves ADHD classification accuracy of 98.30% which is a significant improvement over the state-of-the-art methods.
- The superior performance of ANFIS is further established by comparing its performance to several standard machine learning classifiers like logistic regression, random forest, k-nearest neighbour, and gradient boosting classifier.
- The proposed model also performs exceptionally well on other disease datasets, which confirms its versatility to adapt to different datasets, making it domain independent.

The remaining part of this article is structured as follows: the second section gives a brief description of the related work, the third section provides the details of the dataset and data preprocessing, the fourth section explains the proposed model and the experimental setup, and the fifth section outlines and discusses the experimental results. The last section concludes the article.

## **RELATED WORK**

In 2011, the Neuro Bureau published the ADHD-200 dataset in the public domain to promote research in ADHD diagnosis. Since then, many machine learning models and deep learning models have been proposed for ADHD detection. Personal characteristic data (like age, rating scale scores, IQ scores, gender, handedness, etc.), structural Magnetic Resonance Imaging (sMRI), and functional Magnetic Resonance Imaging (fMRI) scans are the widely used modalities. The ADHD-200 dataset remains the benchmark dataset for ADHD detection in the literature. Brown et al. [29] used a feature subset of the personal characteristic data from the ADHD-200 dataset and fed it to a logistic classifier for ADHD classification. They achieved an accuracy of 62,52%; however, the personal characteristic features were selected randomly, with no selection criteria being specified. S. Ghiassian, et al. [30] proposed an ADHD classification model based on MRI features along with a personal characteristic feature subset that achieved an accuracy of 66.7%. However, the feature subset used in the study was quite small, as they excluded all the features related to the measurement of ADHD symptoms (rating scale scores and IQ scores). Tan et al. [31] used functional brain volume along with personal characteristic features. They developed a Support Vector Machine (SVM) that can distinguish ADHD subjects from non-ADHD individuals. Functional brain volume represented the volumes of those brain regions that were found to be concurrently active during fMRI scanning. This model achieved an improved accuracy of 67,7% but model complexity was very high. Z. Mao, et al. [32] proposed a CNN-LSTM model for ADHD classification. Spatial features were extracted from fMRI using a 3-D CNN, which were subsequently combined with temporal features using a long short-term memory unit. This model achieved an accuracy of

68,8%. A. Riaz, et al. [33] proposed a fusion of imaging and non-imaging data for classifying ADHD. They calculated the functional connectivity of brain regions and integrated it with personal characteristic data to predict the diagnostic label. This model achieved a site-wise accuracy of 86,7%. The mentioned results cannot be generalised as the dataset used for this study was very small. They used data from only 4 sites out of 8 sites. M. Chen, et al. [34] developed a multiscale and multichannel deep neural network(mcdNN) for ADHD classification. It calculated functional connectivity features at different scales (fine, medium, and coarse) using various standard brain parcellation templates. These multiscale brain connectome features, along with personal characteristic features, were used as joint features for ADHD classification. An ADHD classification accuracy of 78,3% was achieved by this model. They also compared the classification accuracy of the model with single-scale brain connectome features and personal characteristic features independently. The model with personal characteristic features alone provided better performance (73,1%) than single-scale brain connectome features (70,6%). This emphasises the importance of personal characteristic features for ADHD classification.

N. Qiang et al. [35] proposed a spatiotemporal attention auto-encoder (STAAE) model to analyse the inputs in spatial and temporal domains simultaneously. An attention score was used to capture the relation b/w different brain volumes in fMRI, and this model achieved a classification accuracy of 72,5%. Yu Huang et al. [36] proposed an ensemble framework called Trans3D-ensemble that combined personal characteristic features with fMRI data for ADHD classification. This model used a 3D transformer to extract spatial features from individual fMRI frames, and a random forest classifier was utilised to extract the personal characteristic features. Both kinds of features were integrated using the stacking ensemble technique and given as input to the final multilayer perceptron model for ADHD classification. This model achieved an accuracy of 74,5%. However, the processing of high spatial dimensions of fMRI images increased the model's complexity manifold. Sudhanshu, et al. [37] proposed a bi-directional long short-term memory (BLSTM) model for ADHD classification. They extracted the time series of voxel activation from 28 different brain regions and gave it to a bidirectional LSTM for ADHD classification. This model achieved a classification accuracy of 87,50% but it cannot be generalised as they used a very small dataset of only 40 subjects. C. Uyulan, et al. [38] used a pre-trained Resnet-50 network for ADHD classification. Using transfer learning, they learned class activation maps for ADHD and healthy children, which provided an accuracy of 93,45 % in ADHD classification.

## **DATA AND PREPROCESSING**

### **DATASET DESCRIPTION**

This article uses personal characteristic data available in the ADHD-200 dataset [13]. It comprises data of 973 participants, with 585 ADHD and 362 non-ADHD participants (for 26 participants, diagnostic prediction is not available). This data was collected by 8 different scanning sites across the world, namely Peking University (PKU), Brown University, Kennedy Krieger Institute (KKI), NeuroIMAGE, New York University Child Study Centre (NYU), Oregon Health and Science University (OHSU), University of Pittsburgh, and Washington University in St. Louis (WashU). The details of various personal characteristic features are provided in Table 1.

**Table 1.** Personal Characteristic features available in the ADHD-200 dataset.

S. No.	Feature Name	Description
1.	Site	Data Collection Site.
2.	Gender	0-Female, 1-Male
3.	Age	Age in years.
4.	Handedness	0-Left Handed, 1-Right Handed
5.	Dx	Diagnostic Label. 0-Non-ADHD 1-ADHD
6.	Secondary Dx	Comments, if any, supporting the diagnosis.
7.	ADHD Measure	Rating scale used to calculate various ADHD scores.
8.	Inattentive	Inattention score of the participant.
9.	Hyper/Impulsive	Hyperactivity score of the participant.
10.	ADHD Index	Combined ADHD score of the participant.
11.	IQ Measure	An intelligence scale used to measure various IQ scores.
12.	Verbal IQ	IQ score of the participant based on acquired knowledge, verbal reasoning, and attention to verbal materials
13.	Performance IQ	IQ score of the participant based on the performance in some visual-spatial tasks, e.g., visual puzzles, picture completion, etc.
14.	Full2 IQ	A subset of the full-scale IQ score based on some memory-related tasks, e.g., number sequence, recalling the photos shown earlier, etc.
15.	Full4 IQ	Another subset of the full-scale IQ score, based on processing speed, e.g., identifying specific symbols/pictures in a row of various symbols/pictures.
16.	Med Status	The medication status of the participant

## DATA PREPROCESSING

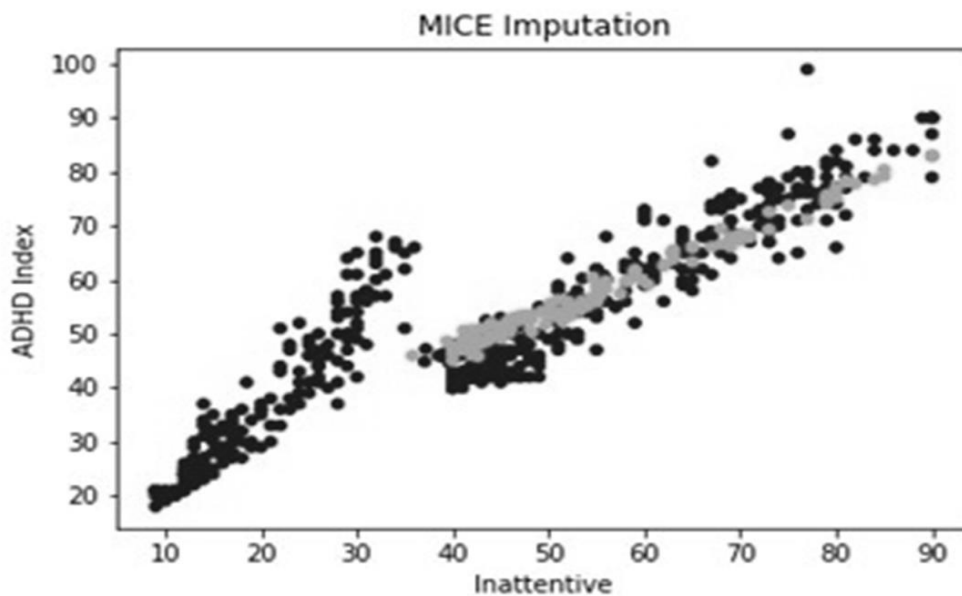
### Data Cleaning

Out of 973 participants, 26 participants, all belonging to Brown University, did not have any diagnostic labels. So, these records are removed from the dataset. 73 participants, all belonging to NeuroIMAGE, and 2 participants belonging to NYU are further excluded as most of the feature values (>70%) are missing for these participants. So, now the remaining number of samples in the dataset is 872, out of which 547 samples are ADHD and 325 samples are non-ADHD. Further details of the data set are shown in Table 2. **Table 2.** Site-wise distribution of the dataset after data cleaning.

Data Collection Site	Total Subjects	ADHD Subjects	Non-ADHD Subjects
Peking University (PKU)	245	143	102
Kennedy Krieger Institute (KKI)	94	69	25
New York University (NYU)	261	110	151
Oregon Health and Science University (OHSU)	113	70	43
University of Pittsburgh	98	94	4
Washington University	61	61	0
<b>Total No. of Subjects</b>	<b>872</b>	<b>547</b>	<b>325</b>

## Data Imputation

Many samples in the dataset still possess missing values for one or the other feature. So, in the next step, these missing values are filled using the MICE technique for imputation [27, 28, 39]. In this technique, each variable with missing data is imputed using the other variables present in the dataset, hence being called the multivariate technique. The default imputation strategy of ‘mean’ is used in the proposed model. Initially, the mean value of the column is substituted in place of the missing values. Next, the imputed values for one variable are set back to missing, and a series of regression models (chained equations) are trained to impute those values using regression [40]. In the next cycle, the value of some other variable is set as missing and is predicted using the regression prediction. This process is repeated until the sequence of the table converges [41]. The scatter plot of imputed vs original data is given in Figure 1.



**Figure 1.** Scatter plot depicting MICE imputation. Black dots represent available data points, and grey dots represent imputed data points.

Figure 1 illustrates the imputation process with the Inattentive score on the x-axis and ADHD-Index on the y-axis. The black point represents the actual data, while grey points indicate the values imputed by MICE for missing values. In this plot, the imputed data points follow the same upward linear trend as observed in actual data, highlighting that MICE preserve the correlation between variables rather than introducing bias. In the above scatter plot, only two variables are plotted for simplifying the visualisation; MICE utilises all available variables in the dataset during the imputation process to ensure robust and consistent estimation.

The current study selects MICE over other simpler methods, such as mean, median, or mode imputation, because it can model the multivariate relationships between variables, as shown in the above scatter plot, and produce multiple plausible imputations rather than a single fixed value [42]. Mean, median, and mode imputation techniques may introduce potential bias, while MICE preserves the underlying correlation structure among features [25], and is more robust for handling missing data.

## PROPOSED MODEL

The proposed model is implemented using Keras with TensorFlow as the back end, and scikit-learn was used to implement the various data imputations, feature selection techniques, and evaluation metrics etc. The workflow of the proposed model is shown in Figure 2. For reproducibility, Algorithm 1 presents the step-by-step implementation of the proposed methodology.

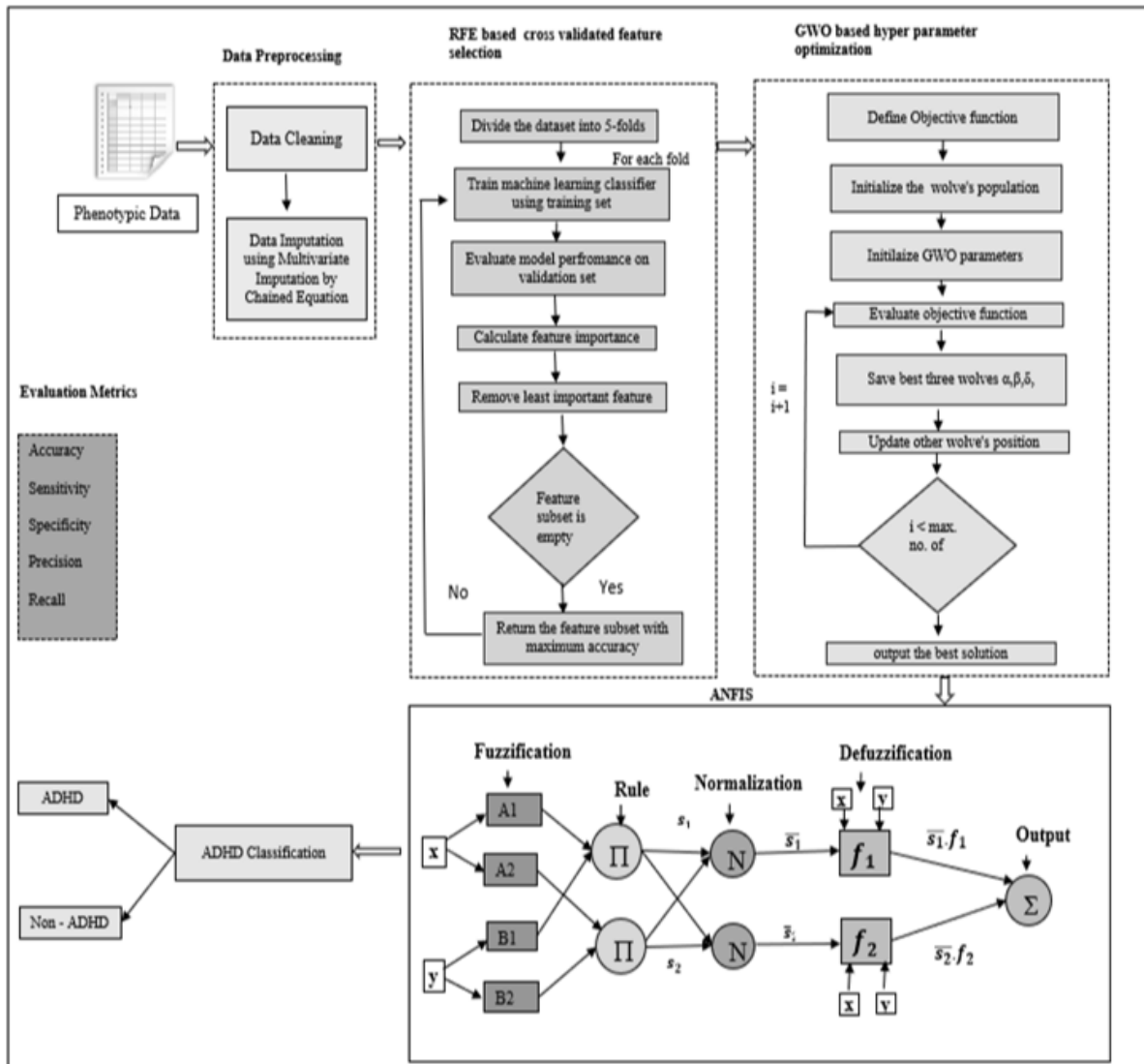


Figure 2. Workflow diagram of the proposed model.

The individual blocks of the proposed model are described further in the text.

### RFE-BASED CROSS-VALIDATED FEATURE SELECTION

RFE[14] is a wrapper-based feature selection technique. It uses a greedy search algorithm to examine various possible feature subsets and selects the one that produces good results for a particular machine learning algorithm. The process of RFE can be shown using the flowchart shown in Figure 3.

The present work employs a decision tree-based classifier as an estimator to assess the importance of the selected features. Another important parameter of RFE is 'n\_features\_to\_select', which requires a specific number of features to select from the feature set. However, it is not always possible to know in advance what number of features will result in the optimal performance of the model. Selecting any random number of features will affect the model's performance. So, to ensure that the optimal number of features is selected, this work integrates a cross-validation technique with RFE known as RFECV.

Cross-validation increases the robustness and generalizability of the model towards unseen data. The value for fold(k) is set to 5, which splits the dataset into 5 equal folds. The decision tree model is repeatedly trained and evaluated five times, each time using a distinct fold as the validation set, and the remaining folds are used as the training set. The mean accuracy across

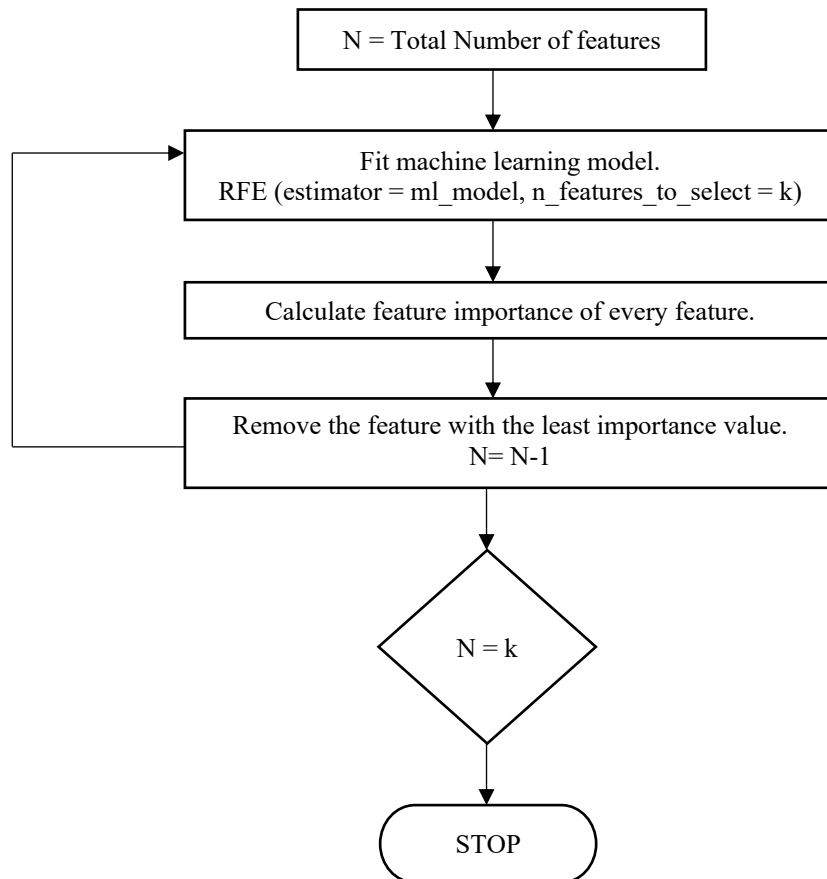
the five different folds is calculated to provide a thorough evaluation of the model's performance. The whole process of RFECV can be summarised using Algorithm 2.

Algorithm 1, highly accurate hybrid model for ADHD classification using ANFIS-RFE-GWO, is as follows:

<b>Input:</b>	ADHD-200 Phenotypic data with missing values.
<b>Output:</b>	(i) Imputed data (ii) Most relevant and optimised feature subset (iii) Hyper-tuned model parameters (iv) Classification labels (ADHD, Normal)
<b>Steps:</b>	
1:	<b>Data Preprocessing</b> (i) Perform data cleaning. (ii) Apply MICE to impute missing values using the 'mean' strategy.
2:	<b>RFE-Based Cross-Validated Feature Selection</b> (i) Divide the dataset into k-folds. (ii) For each fold: Evaluate the feature importance by training and validation on different folds and remove the least important feature. (iii) Repeat until an optimal feature subset is obtained.
3:	<b>GWO-Based Hyperparameter Optimisation</b> (i) Define the objective function (maximise classification accuracy). (ii) Initialise wolves' population and parameters (iii) While iteration < max_iterations: (a) Train the model using the hyperparameters represented by the wolves and evaluate the objective function. (b) Identify $\alpha$ , $\beta$ , $\delta$ wolves. (three best hyperparameter value combinations) (c) Update positions of remaining wolves. (adjust hyperparameter towards better accuracy) (iv) Return the best position of $\alpha$ (best hyperparameters).
4:	<b>ADHD classification using ANFIS</b> (i) Initialise ANFIS hyperparameters with the best values obtained by GWO. (ii) Input the selected feature subset into ANFIS. (iii) Perform fuzzification of inputs. (iv) Apply fuzzy rules generated by ANFIS and normalisation. (v) Perform defuzzification to generate a discrete output. (vi) Obtain Classification labels (ADHD/Normal).
5:	<b>Performance Evaluation</b> (i) Compute Accuracy, Sensitivity, Specificity, Recall, and Precision on the Test dataset.

In the proposed model, RFECV identifies an optimal feature subset containing 10 features. Figure 4a illustrates a plot depicting the relationship between the number of features and the model accuracy.

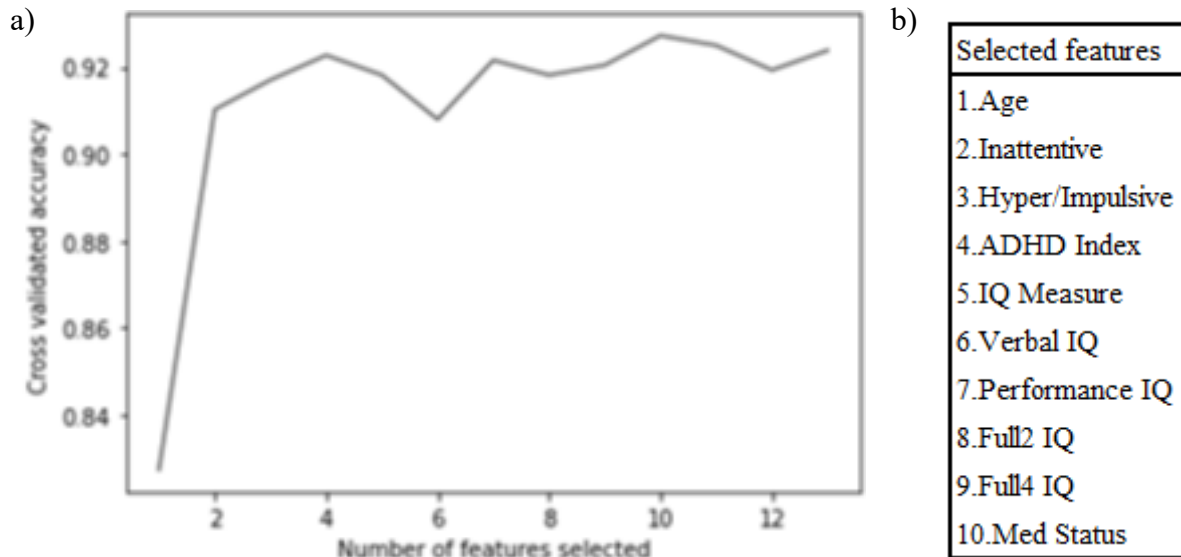
The model's accuracy is the highest when there are 10 features in the selected feature subset. The accuracy decreases if any other features are added or removed. The features selected by RFECV are listed in Figure 4b.



**Figure 3.** Recursive Feature Elimination technique flowchart.

Algorithm 2, recursive feature elimination with cross-validation, is as follows:

- 1: Provide Training dataset T, a Set of 15 features,  $F = \{f_1, f_2, f_3, \dots, f_{15}\}$ , and a Ranking estimator (Decision Tree Classifier) as input.
- 2: **for** all features in F **do**
- 3:     **for**  $k = 1$  to 5 **do**
- 4:         Divide dataset T into five equal parts. One part is used as validation data, and the remaining four parts are used for training purposes.
- 5:         Train the Decision Tree model using the training data.
- 6:         Calculate the classification accuracy of the model using validation data.
- 7:         Obtain the importance of each feature assigned by the model.
- 8:     **end for**
- 9:     Calculate the mean accuracy of the model and the importance of the feature across 5 folds.
- 10:     **If** classification accuracy using the selected feature subset  $F_s$  is highest, **then**
- 11:         Selected features =  $F_s$
- 12:     **else**
- 13:         Drop the least important feature and update the feature subset.
- 14:     **end for**
- 15: Return the best-performing feature subset along with feature rankings.



**Figure 4.** a) Recursive feature elimination with cross-validation, b) features selected by RFECV.

### GWO-BASED HYPERPARAMETER OPTIMIZATION

As mentioned in Section *Major Contributions*, weights, biases, and other network parameters of ANFIS are initialised randomly, potentially leading to convergence towards a local minimum that can impact the overall efficacy of the classification system. To address this issue, the proposed model employs a reliable metaheuristic optimisation technique, GWO [22] that efficiently optimises several hyperparameters of the ANFIS classifier. GWO is based on the leadership hierarchy and hunting behaviour of grey wolves. Grey wolves typically hunt in packs made up of four different hierarchy levels: alpha ( $\alpha$ ), beta ( $\beta$ ), delta ( $\delta$ ), and omega ( $\omega$ ). During hunting, it is assumed that  $\alpha$ ,  $\beta$ , and  $\delta$  have better knowledge of the location of the target prey. So, the position of  $\alpha$  is considered best, the position of  $\beta$  is considered second best, while  $\delta$  is considered at the third best position. These best positions are saved, and other wolves ( $\omega$ ) are forced to update their positions accordingly. This continues till the prey is exhausted or the grey wolves have achieved the best position, after which they attack the prey. The parameter setting of GWO is described in Table 3.

**Table 3.** Parameter setting for grey wolf optimisation.

Parameter Name	Parameter Value
No. of wolves	10
No. of iterations	20
Members in the parameter vector	7
Search domain	Same as the range specified for individual hyperparameters in Table 4.

In the optimisation context, the targeted prey is the optimised value of the hyperparameters, which is achieved using the following steps:

- (i) Objective Function ( $O_f$ ): GWO investigates various hyperparameter combinations against some predefined criteria. Therefore, it is necessary to define an objective function to assess these combinations. The main objective of any optimisation process is to maximise classification accuracy. So, the objective function can be defined as follows:

$$O_f = \max \left( \text{acc} \left( \text{ANFIS}(\overline{\text{HP}}) \right) \right) \quad (1)$$

where  $\overline{\text{HP}}$  represents the hyperparameter vector of the model.

- (ii) Population Initialisation: The population size is initialised as 10. Each wolf is referred to as an agent. The starting position of every wolf is initialised at random from the range specified for different hyperparameters, as mentioned in Table 4.
- (iii) Fitness Evaluation: In this step, GWO is trained, and each agent position (i.e., hyperparameter values) is evaluated corresponding to the fitness value (classification accuracy), and the best top three positions –  $\alpha$ ,  $\beta$ , and  $\delta$  – are selected. These positions are used to update the positions of other agents.
- (iv) Iterative Optimisation: This process of fitness evaluation, choosing the best three solutions, and updating the position is repeated for the specified number of iterations or till the optimisation process converges. Finally, the best value of hyperparameters ( $\alpha$ ) resulting in maximum accuracy is chosen for training the ANFIS model.

In the present study, ANFIS hyperparameters are optimised by first considering the input membership function type (triangular-shaped, generalised bell-shaped, and Gaussian), and the number of membership functions (in the range of 1 to 8). Then, in the next stage, weights and biases of different ANFIS nodes, number of training epochs, learning rate, and batch size for the training process are optimised to achieve the most optimal model performance. The details of various hyperparameters, along with their possible values and optimised values provided by GWO, are provided in Table 4.

**Table 4.** Hyperparameter details along with optimised values.

Hyperparameter Name	Possible values	Optimised values
Membership function	Triangular-shaped, Generalised bell-shaped, Gaussian-shaped	Gaussian-shaped
Number of membership functions	from 1 to 8	3
Weights	from -2 to 2	Differ for different nodes
Biases	from -2 to 2	Differ for different nodes
Number of training epochs	10 to 100	50
Learning Rate	$10^{-1}$ , $10^{-2}$ , $10^{-3}$ , $10^{-4}$	$10^{-3}$
Batch Size	8, 16, 32, 64, 128	16

### ADHD CLASSIFICATION USING ANFIS

After the optimisation of the model hyperparameters, the next step involves classification. To perform the classification, the dataset is split into training and test datasets using a ratio of 80:20. All the selected features are normalised using the following formula:

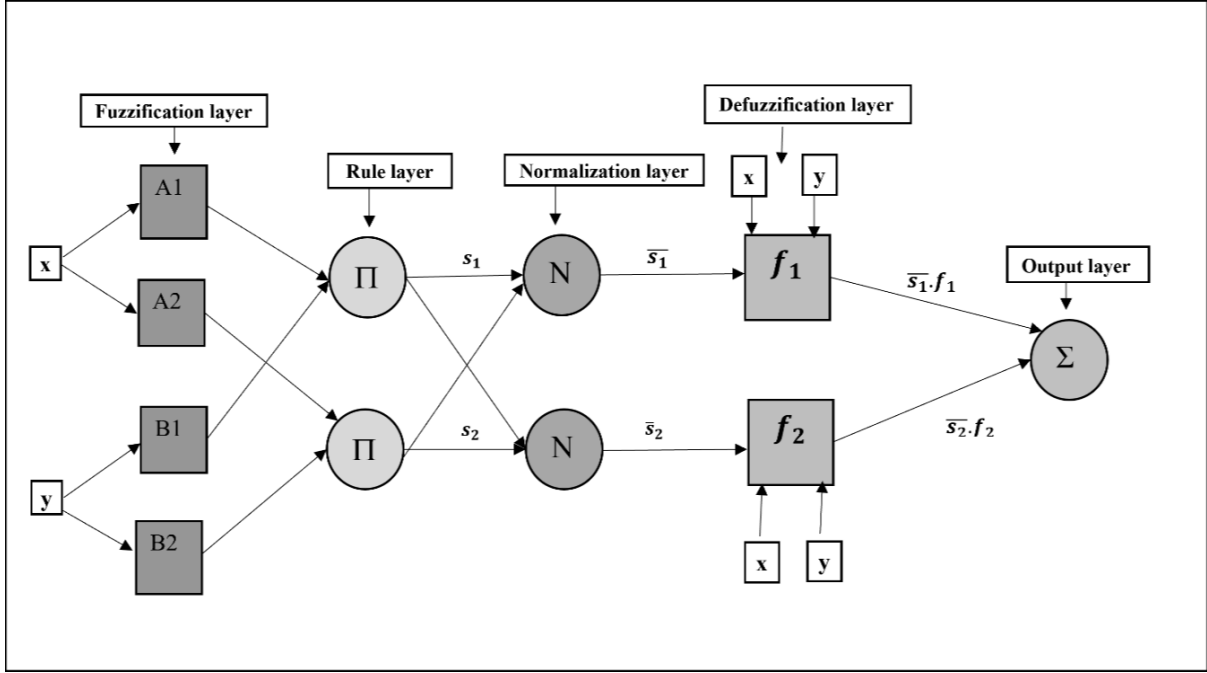
$$\hat{x} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

Normalisation of the data helps in reducing the inconsistencies present within the dataset and mitigating the effect of noise and outliers. Subsequently, the normalised data is fed into ANFIS for classification, whose parameters are already optimised using GWO.

ANFIS can be described as a fuzzy inference system (FIS), integrated into the adaptive framework of an artificial neural network (ANN). It is a 5-layered architecture, as illustrated in Figure 5. The description of each layer is as follows:

- (i) *Fuzzification layer* – there are 10 nodes in this layer since RFECV selected 10 features. Each node is an adaptive node represented by a square. Adaptive nodes are trainable, and their value can be tuned during the training process. This layer receives the non-fuzzy (crisp) inputs and converts them to a set of fuzzy values by applying the membership functions and determining their degree of membership according to equation (3)

$$O_{i1} = \mu_{A_i}(x), \tag{3}$$



**Figure 5.** The structure of a 2-input-based adaptive neuro-fuzzy inference system (ANFIS).

where  $x$  is the input to the node  $i$ ,  $A_i$  refers to quantifiers associated with the membership function  $\mu$  and  $O_{i1}$  refers to the output of the first layer for the node  $i$ .

In the proposed model, the Gaussian function (selected by GWO) is implemented as the membership function. The associated degree of membership function is calculated for each input and forwarded to the rule layer.

- (ii) *Rule Layer* – this layer consists of fixed nodes, i.e., non-trainable nodes, which are represented by a circle. Each node contains fuzzy sets of rules for the input data. The grid partitioning method is used for rule generation. The input data space is divided into arbitrary partitions called grids, and the inference result for each one of those regions is represented in the form of rules. The firing strength (weight) of each rule is calculated using membership values received from the first layer, as shown in equation (4).

$$O_{i2} = s_i = \mu_{A_i}(x) \cdot \mu_{B_i}(y), \quad i = 1, 2, \quad (4)$$

where  $O_{i2}$  refers to the output of the second layer  $s_i$  refers to the firing strength of the particular rule corresponding to the node  $i$ ,  $\mu_{A_i}$  and  $\mu_{B_i}$  represent the degree of membership of the input  $x$  and  $y$  corresponding to quantifiers  $A_i$  and  $B_i$  received from the first layer.

- (iii) *Normalisation Layer* – this layer calculates the normalised firing strength of each rule. It is done by dividing the firing strength of an individual rule by the total firing strengths of all the rules according to equation (5).

$$O_{i3} = \bar{s}_i = \frac{s_i}{\sum_i s_i}, \quad i = 1, 2, \quad (5)$$

where  $O_{i3}$  refers to the output,  $s_i, \bar{s}_i$  refer to the firing strength and the normalised firing strength of individual rules, respectively. Each node in this layer is non-trainable and therefore represented by a circle.

- (iv) *Defuzzification Layer* – each node in this layer is an adaptive(trainable) node, represented by a square. This layer transforms the fuzzy inference results into a crisp *output*. Takagi-Sugeno approach [43] is used for this purpose. This approach uses the normalised firing strength along with the original input values to calculate the crisp output value as shown in equation (6).

$$\mathbf{O}_{i4} = \bar{s}_i \cdot \mathbf{f}_i = \bar{s}_i(\mathbf{p}_i \mathbf{x} + \mathbf{q}_i \mathbf{y} + \mathbf{r}_i), \quad i = 1, 2, \quad (6)$$

where  $\mathbf{O}_{i4}$  refers to the output of the fourth layer,  $\{\mathbf{p}_i, \mathbf{q}_i\}$  are the weights and  $\mathbf{r}_i$  is the bias corresponding to the inputs for the node  $i$ . The required weights and biases are initialised using Keras random initialisers, which are subsequently optimised using GWO.

- (v) *Output layer* – this layer contains a single fixed node, represented by a circle. It produces the final output by calculating the summation of all the outputs generated by the previous layer according to equation (7)

$$\mathbf{O}_{i5} = \sum_i \bar{s}_i \cdot \mathbf{f}_i = \frac{\sum_i s_i \cdot \mathbf{f}_i}{\sum_i s_i}, \quad (7)$$

where  $\mathbf{O}_{i5}$  refers to the output of the final layer,  $\bar{s}_i$  represents the normalised firing strength and  $\mathbf{f}_i$  represents the linear function calculated in the previous layer for the node  $i$ . This output value is passed to the sigmoid function to get the associated binary classification label.

## TRAINING OF ANFIS MODEL

ANFIS uses a hybrid learning strategy [44, 45] for training that combines the least squares method with the backpropagation technique to train the model. The model is trained in two passes: the forward pass and backward pass. In the forward pass, the least squares method is used to calculate the error for the trainable parameters. In the backward pass, the gradient descent method is used to backpropagate the error rates toward the input layer and update the model parameters accordingly. To enhance the robustness of the model, the training set is further divided into training and validation sets using 5-fold cross-validation. One subset is used as a validation set while the other four are treated as the training set. ANFIS is trained repeatedly until each fold has been treated as a validation set. A learning rate of  $10^{-3}$  and a batch size of 16 (provided by GWO) are used for the training process. The model is trained for 50 epochs to achieve reliable results.

## RESULTS AND DISCUSSION

To determine how effectively the proposed model generalises unseen data, its performance is evaluated on test data using different evaluation metrics as mentioned further in the text.

### EVALUATION METRICS

- (i) *Accuracy* – accuracy (ACC) is a fundamental metric that assesses how well the model predicts an instance class in the test set. Accuracy is calculated as:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}, \quad (8)$$

where TP (True Positive) is number of samples correctly classified as ADHD, TN (True Negative) number of samples correctly classified as non-ADHD, FP (False Positive) number of normal samples, wrongly classified as ADHD and FN (False Negative) number of ADHD samples, wrongly classified as non-ADHD.

- (ii) *Sensitivity* – it measures the model’s ability to efficiently identify all the ADHD samples. A high rate of sensitivity means that there are very few unidentified cases of ADHD. It is calculated as follows:

$$\text{Sensitivity} = \frac{TP}{TP+FN}, \quad (9)$$

- (iii) *Specificity* – it reflects the model’s ability to efficiently identify the non-ADHD cases and is calculated as:

$$\text{Specificity} = \frac{TN}{TN+FP}, \quad (10)$$

(iv) *Precision* – precision is calculated to determine, out of the total positive predictions, how many samples are positive. It is calculated as:

$$\mathbf{Precision} = \frac{TP}{TP+FP}, \quad (11)$$

(v) *Recall* – it determines how many positive samples are identified correctly. It is like Sensitivity and calculated as:

$$\mathbf{Recall} = \frac{TP}{TP+FN}, \quad (12)$$

(vi) *F1-Score* – it is an overall measure of the model's performance that combines the precision and recall of the classifier and is calculated as:

$$\mathbf{F1-Score} = 2 \left( \frac{\mathbf{Precision} \cdot \mathbf{Recall}}{\mathbf{Precision} + \mathbf{Recall}} \right), \quad (13)$$

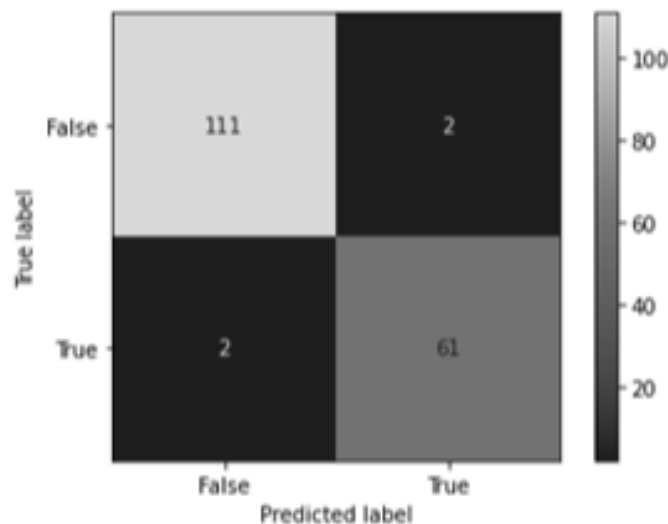
## EVALUATION RESULTS

The proposed model achieves a binary classification accuracy of 99,62% on the training set and an accuracy of 98,30% on the test dataset. The results of various evaluation matrices on test data are shown in Table 5.

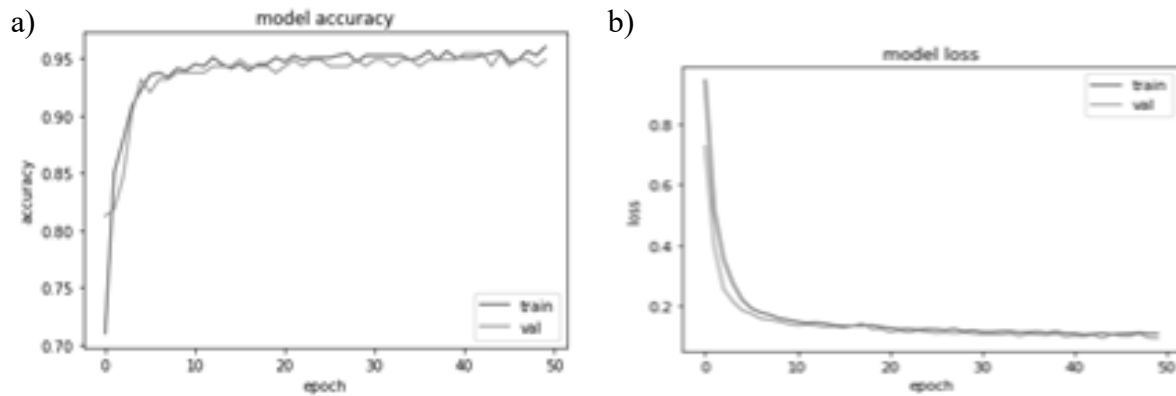
**Table 5.** Evaluation results.

Performance matrices on the test dataset			
Accuracy	98,30%	Precision	98,38%
Sensitivity	96,82%	Recall	96,82%
Specificity	96,02%	F1-Score	97,60%

Another quantitative assessment of the proposed model is done using the confusion matrix. It is quite useful for determining which classes are being incorrectly classified frequently. The confusion matrix of the proposed model is shown in Figure 6. The proposed model misclassifies only 4 out of 176 test samples, showing the superiority of the proposed approach. Since the input dataset is imbalanced, i.e. one class has more samples than the other class. So, it is important to consider some other evaluation metrics that provide a balanced assessment of the model's performance. F1-score is very useful in such cases, as it takes both the false positives and false negatives into consideration. The proposed model achieves an F1-score of 97,60%, reaffirming the robustness of the proposed model. In addition to quantitative assessment, model performance is evaluated qualitatively using accuracy/loss graphs against the number of epochs. The training accuracy and loss values per epoch are plotted in Figure 7a) and Figure 7b), respectively.



**Figure 6.** Confusion matrix for test dataset.



**Figure 7.** a) Training Accuracy curve w.r.t epoch, b) Training Loss curve w.r.t epoch.

### Comparison with Different Machine Learning Classifiers

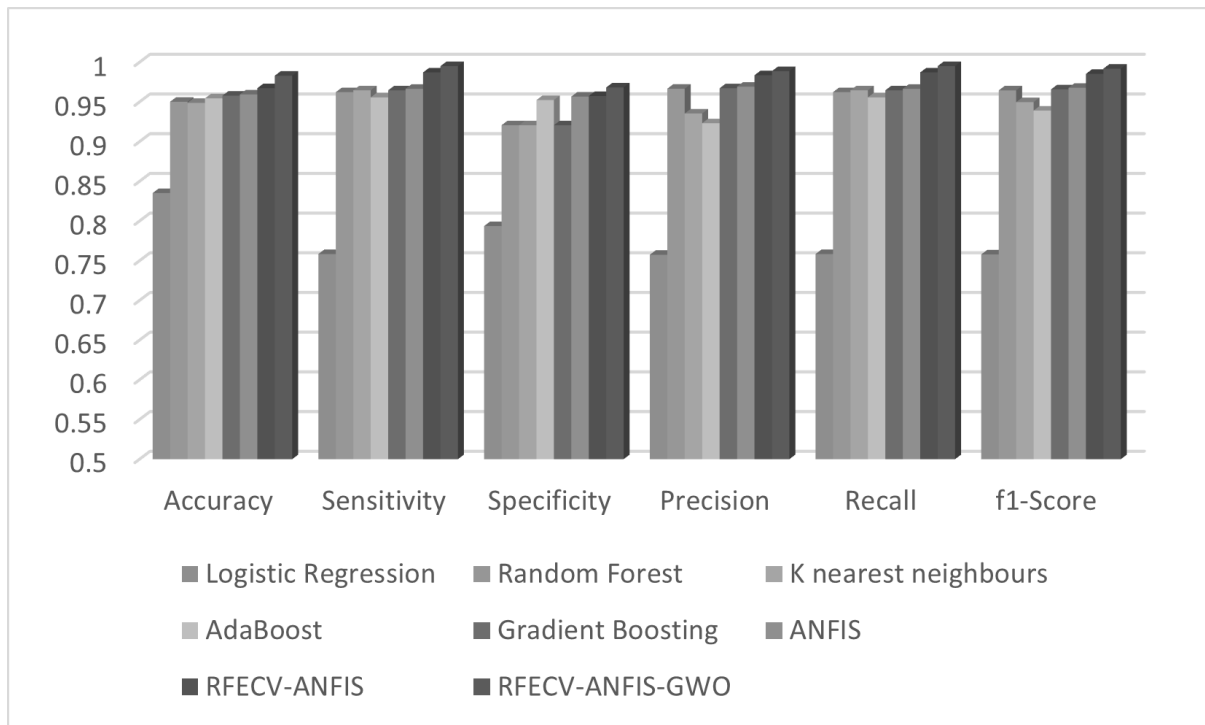
The performance of the proposed model is compared with other machine learning classifiers, such as Logistic Regression, Random Forest, K-Nearest Neighbour, AdaBoost, and Gradient Boost classifier, to evaluate the efficacy of the model. The model's performance is also compared using the ANFIS classifier in the absence of feature selection and optimisation. The comparison results are shown in Table 6.

**Table 6.** Performance comparisons of different ML classifiers on the test dataset.

Classifier	Accuracy	Sensitivity	Specificity	Precision	Recall	f1-score
Logistic Regression	0,8352	0,7584	0,7936	0,7575	0,7584	0,7579
Random Forest	0,9502	0,9623	0,9206	0,9666	0,9623	0,9645
K nearest neighbours	0,9488	0,9646	0,9206	0,9354	0,9646	0,9497
AdaBoost	0,9545	0,9557	0,9523	0,9231	0,9557	0,9391
Gradient Boosting	0,9579	0,9646	0,9206	0,9672	0,9646	0,9658
ANFIS	0,9594	0,9665	0,9567	0,9692	0,9665	0,9678
RFECV-ANFIS	0,9673	0,9671	0,9574	0,9836	0,9671	0,9753
<b>Proposed Method</b>	<b>0,9830</b>	<b>0,9682</b>	<b>0,9602</b>	<b>0,9838</b>	<b>0,9682</b>	<b>0,9760</b>

The performance comparison described above is shown graphically in Figure 8.

From Table 6 and Figure 8, it can be inferred that the ANFIS classifier performs better than other machine learning classifiers and achieves an accuracy of 95,94%. This is because of the efficient handling of data uncertainties by ANFIS, employing a human-like reasoning approach through fuzzy rules. Also, ANFIS is much simpler and more effective than general fuzzy inference systems, as it does not require prior domain expertise for rule formation. The proven learning capabilities of neural networks are utilised for defining the rules. The effect of applying feature selection is also inspected for the proposed model. It is evident from Table 6 that with the inclusion of the feature selection technique (RFECV), the classification accuracy increased to 96,73%. Similarly, there is an improvement in other performance metrics as well. This improvement in model performance emphasises how crucial it is to choose an effective feature subset for optimal model performance. The classification performance experiences a significant improvement with the optimisation of hyperparameters through the grey wolf optimisation technique. The proposed model attains the highest accuracy of 98,30% by integrating ANFIS with GWO, validating the use of optimisation techniques.



**Figure 8.** Graphical representation of performance comparison between the proposed model and other ML classifiers on test data.

### Comparison with Existing Methods

To further demonstrate its efficiency, the proposed model is compared against some of the state-of-the-art classifiers in ADHD diagnosis. The comparison results are shown in Table 7.

**Table 7.** Performance comparisons of different models on the ADHD-200 dataset.

Method	Accuracy, %
CNN- LSTM [32]	68,8
STAAE [35]	72,50
mcDNN [34]	78,3
A. Riaz et al. [33]	86,7
BLSTM [37]	87,50
Uyulan C et al. [38]	93,45
<b>Proposed Method</b>	<b>98,30</b>

### Algorithm Complexity

The computational complexity of the proposed framework arises primarily from MICE imputation, RFECV feature selection, and GWO-based hyperparameter optimisation. MICE requires iterative regression across variables, with a complexity of  $O(N \cdot d^2)$  where ‘ $N$ ’ is the no. of samples and ‘ $d$ ’ is the number of features. RFECV has a complexity of  $O(k \cdot d \cdot T)$  with  $k$ -fold cross-validation and ‘ $T$ ’ is the training time of the classifier. GWO complexity depends on the number of wolves ( $P$ ) and iterations ( $I$ ), with complexity  $O(P \cdot I \cdot T)$ . ANFIS is relatively efficient with complexity linear in the number of rules and inputs ( $O(R \cdot d)$ ).

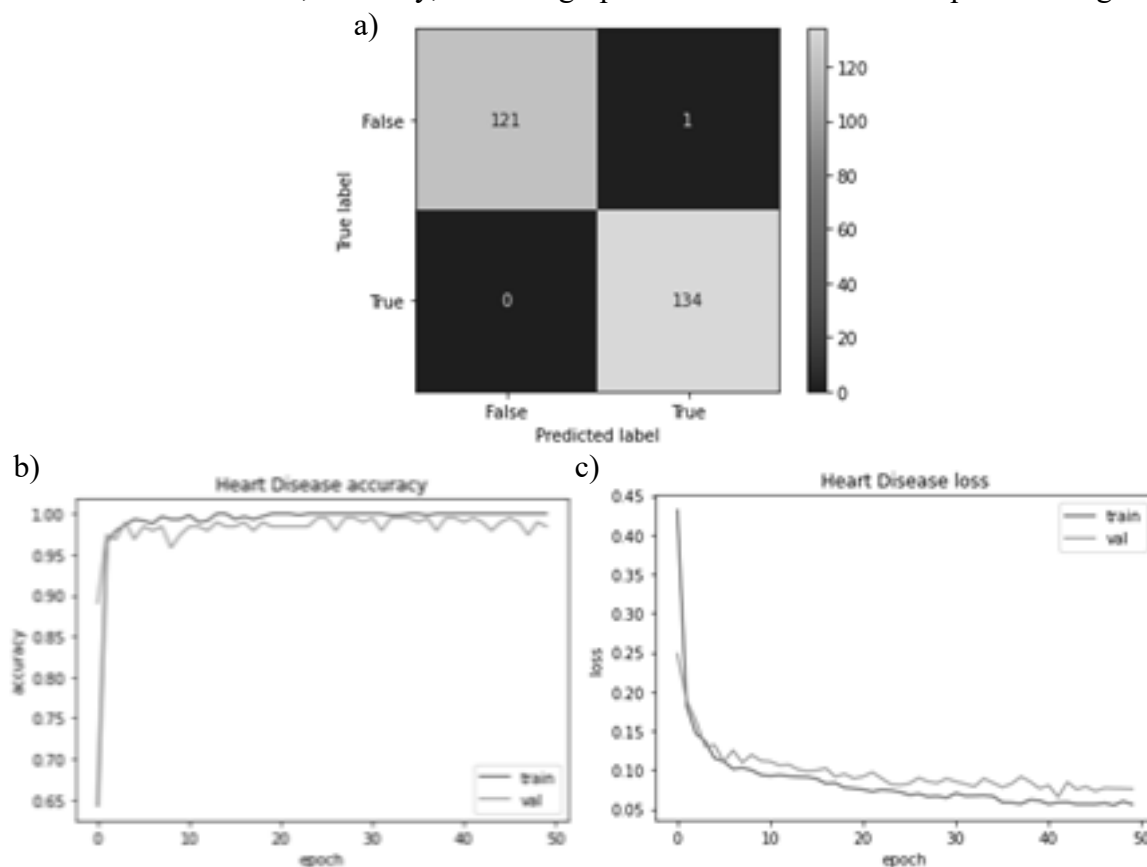
Despite the multiple stages, the framework remains computationally feasible and suitable for practical applications due to dimensionality reduction (RFECV), efficient metaheuristic convergence (GWO), and the relatively lightweight structure of ANFIS compared to deep learning models.

## VALIDATION OF RESULTS ON OTHER DATASETS

To assess its applicability in different domains, the proposed model is evaluated on four benchmark disease datasets, including the Cleveland Heart Disease dataset, Chronic Kidney Disease dataset, COVID-19 dataset, and Breast Cancer Wisconsin dataset. These datasets vary in the number of instances and features they contain. Some datasets have fewer features but a larger number of instances, while others have more features but fewer instances. This is done to assess the model's performance across diverse dataset features. The following sections further elaborate on the evaluation results and specific details regarding the datasets.

### Cleveland Heart Disease Dataset

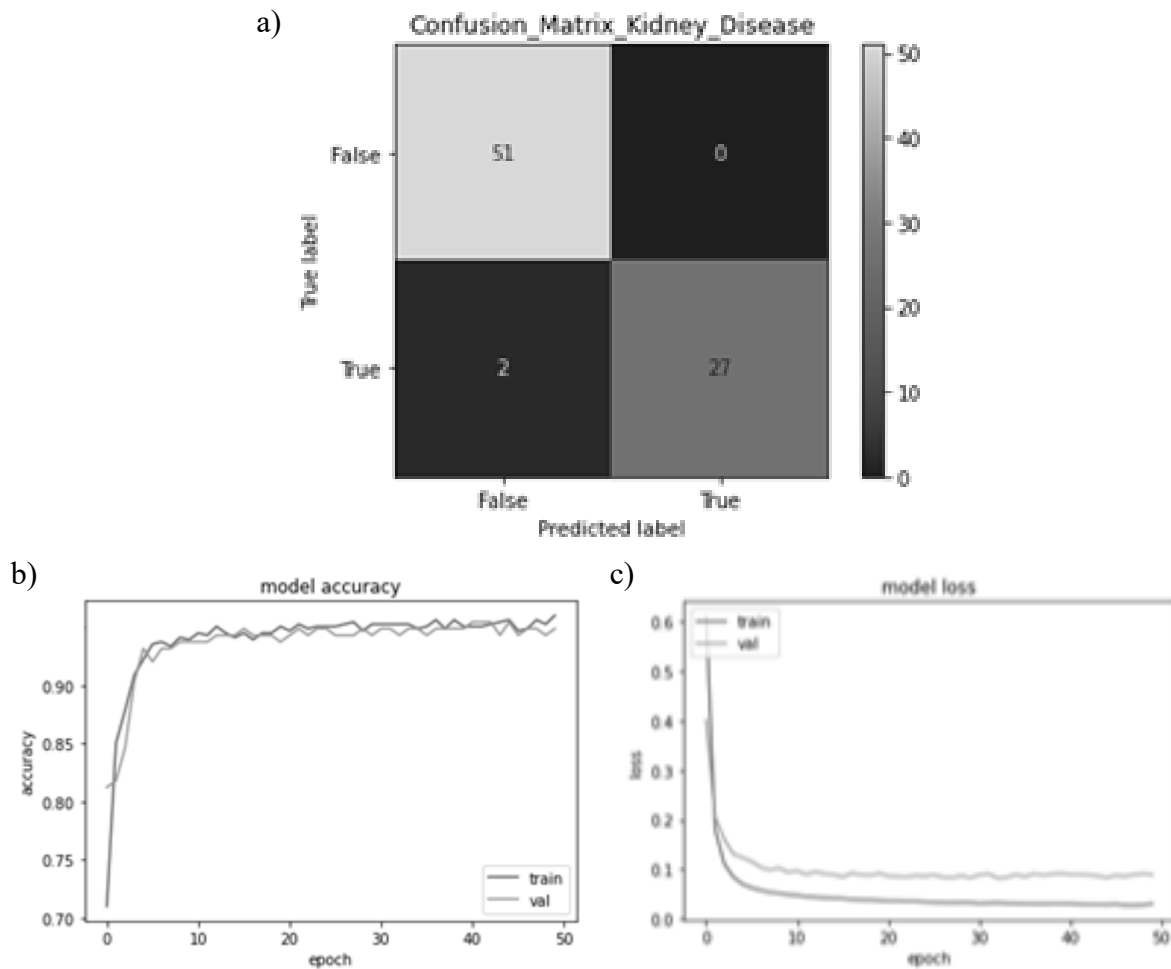
The Cleveland Heart Disease dataset [46] is the benchmark dataset used for heart disease prediction. It is obtained from the Kaggle Machine Learning repository. It comprises data from 1025 individuals collected from Cleveland, Hungary, Switzerland, and the VA Long Beach. The dataset includes 13 input features like age, blood pressure, chest pain experienced, cholesterol level, blood sugar, and maximum heart rate, etc. along with a target label. The model's performance is assessed using this dataset, with all the experimental settings being consistent as mentioned in the Section *Proposed Model*. Out of 13 input features, 8 features are selected based on their significance through the feature selection technique (RFECV), which are further used by ANFIS for classification. The proposed model achieves an outstanding accuracy of 100% on the training dataset while maintaining a similar high *Accuracy* of 99,32% on the test dataset. The proposed model also demonstrates exceptional performance corresponding to various evaluation measures, achieving a *Sensitivity* value of 100%, a *Specificity* of 98,18% a *Precision* of 98,55%, and an *F1-score* of 99,27%. The outcomes of the confusion matrix, accuracy, and loss graph for the evaluation are depicted in Figure 9.



**Figure 9.** a) Confusion matrix for the Cleveland Heart Disease test dataset, b) Training Accuracy curve and c) Training Loss curve w.r.t epoch for the Cleveland Heart Disease dataset.

## Chronic Kidney Disease Dataset

The Chronic Kidney Dataset [47] is the standard dataset for kidney disease prediction. It is taken from the UCI machine learning repository. It comprises data from 400 individuals, including 24 input features like age, blood pressure, sugar level, red blood cell count, white blood cell count, sodium level, potassium level, hypertension, etc., along with a target label. 14 input features are selected through the feature selection technique, which contributes most towards the classification result. The model proposed in this study attains a remarkably high accuracy of 99,24% on the training set while achieving an *Accuracy* of 97,15% on the test dataset, which aligns with leading research in the field. The model exhibits strong performance across various evaluation measures, with a *Specificity* of 95,10% and a *Precision* of 100%, a *Sensitivity* of 92,90%, and an *F1-score* of 96,32%. The assessment results, including the confusion matrix, accuracy, and loss graph, are illustrated in Figure 10.

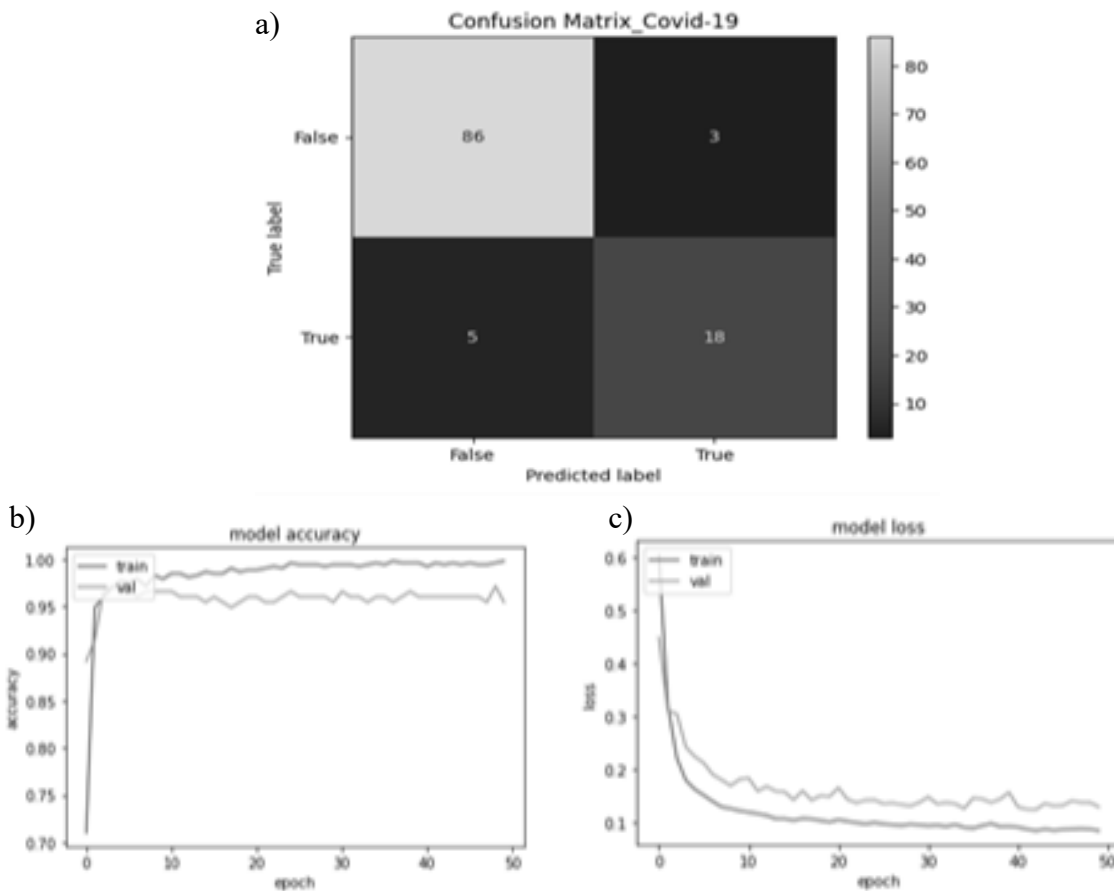


**Figure 10.** a) Confusion matrix, b) training accuracy curve and c) training loss curve w.r.t epoch, all for chronic kidney disease dataset.

## Covid-19 Dataset

The COVID-19 Dataset [48] is publicly available and provided by Alakus and Turkoglu. The dataset includes 600 instances and 18 clinical features such as platelets, red blood cells, haemoglobin, serum glucose, haematocrit, urea, creatinine, etc., along with a diagnostic label. The feature selection technique selects the 10 most relevant features, which are further utilised by ANFIS for classification purposes. The proposed model attains an outstanding accuracy of 99,22% on the training dataset while simultaneously maintaining a comparable high *Accuracy*

of 95,83% on the test dataset. The proposed model also showcases impressive performance across various evaluation measures, achieving *Specificity* of 96,34%, *Sensitivity* of 87,96%, a *Precision* of 85,27%, and an *F1-score* of 86,59%. The results of the confusion matrix, accuracy, and loss graph for the evaluation are illustrated in Figure 11(a-c).



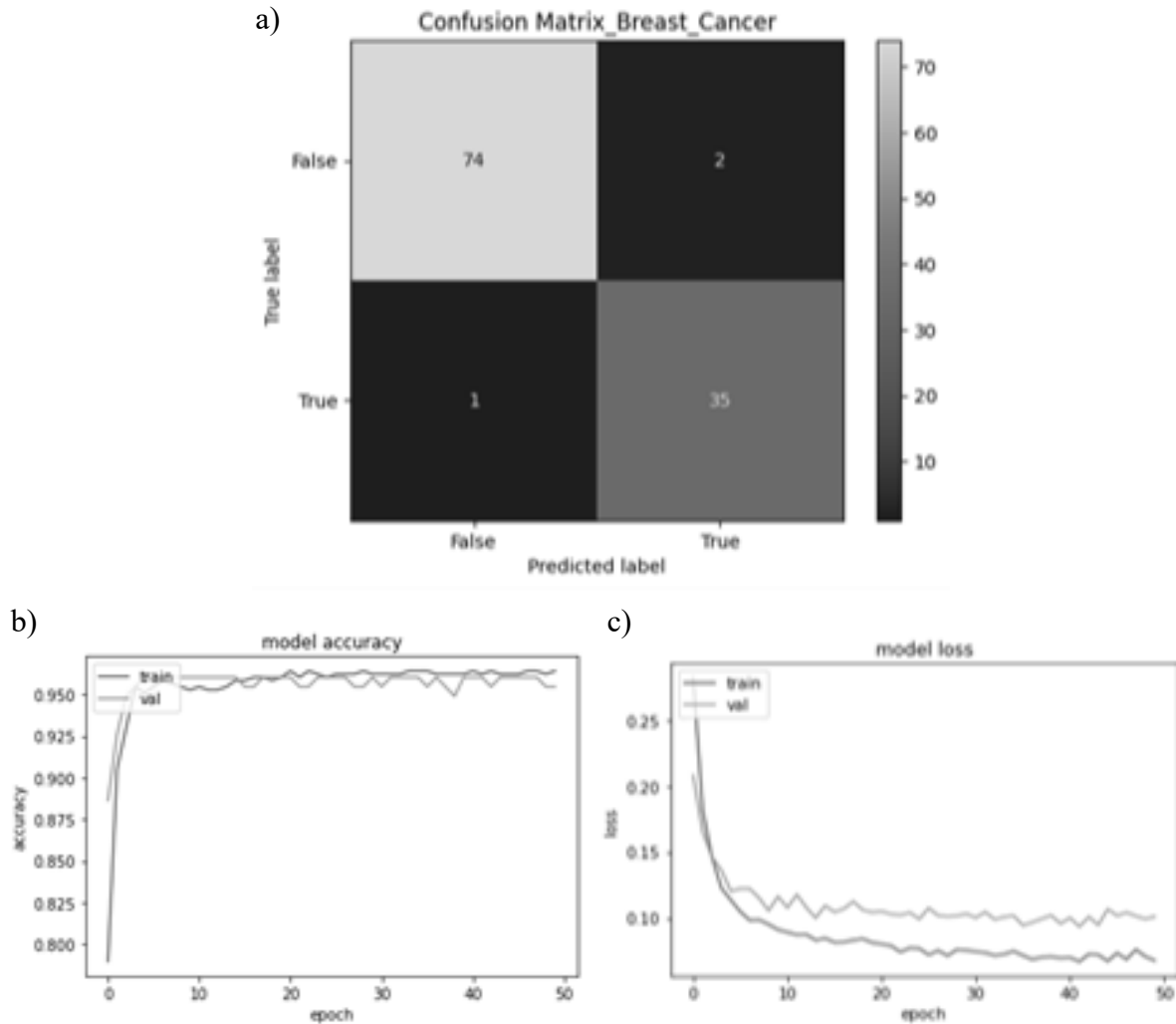
**Figure 11.** a) Confusion matrix for the COVID-19 test dataset, b) training accuracy curve and c) training loss curve w.r.t epoch, all for COVID-19 dataset.

### Breast Cancer Wisconsin Dataset

The Breast Cancer Wisconsin Dataset [49] is the benchmark dataset for breast cancer prediction. It is taken from the UCI machine learning repository. It includes 569 instances featuring 29 distinct input features such as radius, perimeter, texture, smoothness, compactness, concavity, and area, among others, in addition to a target label. Twelve input features are chosen utilising the feature selection method that has the greatest impact on the outcome of the classification process. The proposed model achieves a very high accuracy of 98,54% on the training set while achieving an *Accuracy* of 96,73% on the test dataset, which is comparable with state-of-the-art research. The model demonstrates robust performance across various evaluation measures with *Specificity* at 96,64% and *Sensitivity* at 97,22%, *Precision* at 93,65%, and *F1-score* of 95,40%. The assessment results, including the confusion matrix, accuracy, and loss graph, are illustrated in Figure 12.

The evaluation results on different disease datasets are summarised in Table 8.

The results from Table 8 illustrate the exceptional and consistent performance of the proposed model across diverse sets of disease data, each distinguished by a unique set of features. This signifies the model's versatility and robustness irrespective of the specific domain, thus making it an indispensable tool for clinical decision-making.



**Figure 12.** a) confusion matrix, b) training accuracy curve and c) training loss curve w.r.t epoch, all for Breast Cancer Wisconsin dataset.

**Table 8.** Evaluation results on various benchmark disease datasets.

Classifier	Training Accuracy	Test Accuracy	Sensitivity	Specificity	Precision	f1-Score
Cleveland Heart Disease Dataset	1,0000	0,9932	1,0000	0,9818	0,9855	0,9927
Chronic Kidney Disease Dataset	0,9924	0,9715	0,9290	0,9510	1,0000	0,9632
COVID-19 Dataset	0,9922	0,9583	0,8796	0,9634	0,8527	0,8659
Breast Cancer Wisconsin Dataset	0,9854	0,9673	0,9722	0,9664	0,9365	0,9540

### LIMITATIONS AND FUTURE WORK

There are certain constraints associated with the proposed model that need to be considered as well. The model is evaluated on the datasets having (~1k) instances. So, it needs to be applied to some extensive datasets(>1k) to further strengthen the model's outcomes. While ADHD-200 and other mentioned datasets contain diagnostic information about the disease subtypes as well, this work focuses only on the binary classification of ADHD and other disorders. So, the multiclass classification remains in the future scope of this work.

## CONCLUSION

This article proposes a highly accurate, reliable, efficient, and versatile hybrid model based on metaheuristic and fuzzy logic for ADHD classification. The primary focus lies in the efficient handling of the data uncertainty inherent in medical data. The proposed model leverages the proven inferencing capabilities of fuzzy logic, integrating it with the remarkable learning capacity of neural networks, specifically known as ANFIS. To mitigate the impact of noise and non-significant features, an efficient feature selection technique, recursive feature elimination (RFE), is utilised, which selects the most significant and discriminative features. RFE is further integrated with cross-validation to ensure the robustness of the selected features, which are then utilised by the ANFIS classifier for ADHD classification. The model performance is highly dependent on its hyperparameters, so to counteract their impact, an efficient and reliable meta-heuristic optimisation technique, GWO, is employed by the proposed system.

The empirical findings on the phenotypic data of the ADHD-200 dataset indicate that the ANFIS classifier outperforms various other machine learning classifiers, such as the Random Forest, K-Nearest Neighbours, AdaBoost, and Gradient Boosting classifiers, etc., even in the absence of any feature selection process. Employing RFE for feature selection not only reduces model complexity and computational time but also significantly enhances classification accuracy. Further optimisation of the model's hyperparameters using GWO leads to a substantial improvement in classification performance. The proposed hybrid model achieves an outstanding accuracy of 98.30% which is superior to previous state-of-the-art approaches in ADHD classification. Both quantitative and qualitative evaluations on different performance metrics reaffirm the robustness of the proposed approach. To test the robustness of the proposed approach, the model is evaluated across multiple benchmark disease datasets. These datasets differ in the nature and dimensionality of the features. The exceptional performance of the proposed model on a varied set of data strengthens our belief in the versatility and adaptability of the proposed model to a broad spectrum of medical domains.

An additional significant contribution of this study is the prioritisation of personal characteristic data over MRI scans for ADHD classification. Personal characteristic data possesses a great potential for ADHD diagnosis. In the original ADHD-200 Global Competition, too, a straightforward personal characteristic-only approach [29] surpassed various imaging-based diagnostic techniques. The proposed work demonstrates a way of enhancing such personal characteristic data-only approaches and offers a fresh viewpoint for subsequent research on ADHD classification, aiding healthcare professionals in diagnosing the disorder at an early stage. As a future research plan, it is possible to implement the proposed model for ADHD subtype classification as well.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

The dataset used in the present work is publicly available on the link [http://www.nitrc.org/frs/?group\\_id=383](http://www.nitrc.org/frs/?group_id=383).

## REFERENCES

- [1] Polanczyk, G. and Jensen, P.: *Epidemiologic Considerations in Attention Deficit Hyperactivity Disorder: A Review and Update*. Child and Adolescent Psychiatric Clinics of North America **17**(2), 245-260, 2008, <http://dx.doi.org/10.1016/j.chc.2007.11.006>,
- [2] Danielson, M.L., et al.: *Prevalence of Parent-Reported ADHD Diagnosis and Associated Treatment Among U.S. Children and Adolescents, 2016*. Journal of Clinical Child & Adolescent Psychology **47**, 199-212, 2018, <http://dx.doi.org/10.1080/15374416.2017.1417860>,

- [3] Song, P., et al.: *The Prevalence of Adult Attention-Deficit Hyperactivity Disorder: A Global Systematic Review and Meta-Analysis*.  
Journal of Global Health **11**, No. 04009, 2021,  
<http://dx.doi.org/10.7189/jogh.11.04009>,
- [4] Li, Y., et al.: *Prevalence and Trends in Diagnosed ADHD Among US Children and Adolescents, 2017-2022*.  
JAMA Network Open **6**(10), No. e2336872,2023,  
<http://dx.doi.org/10.1001/jamanetworkopen.2023.36872>,
- [5] Ma, K., et al.: *Functional Overlaps Exist in Neurological and Psychiatric Disorders: A Proof from Brain Network Analysis*.  
Neuroscience **425**, 39-48, 2020,  
<http://dx.doi.org/10.1016/j.neuroscience.2019.11.018>,
- [6] Shen, C., et al.: *What Is the Link Between Attention-Deficit/Hyperactivity Disorder and Sleep Disturbance? A Multimodal Examination of Longitudinal Relationships and Brain Structure Using Large-Scale Population-Based Cohorts*.  
Biological Psychiatry **88**(6), 459-469, 2020,  
<http://dx.doi.org/10.1016/j.biopsych.2020.03.010>,
- [7] Vilar-Ribó, L., et al.: *Genetic Overlap and Causality between Substance Use Disorder and attention-deficit and Hyperactivity Disorder*.  
American Journal of Medical Genetics Part B: Neuropsychiatric Genetics **186**(3), 140-150,2021,  
<http://dx.doi.org/10.1002/ajmg.b.32827>,
- [8] Schein, J., et al.: *Economic Burden of Attention-Deficit/Hyperactivity Disorder among Adults in the United States: A Societal Perspective*.  
Journal of Managed Care & Specialty Pharmacy **28**(2), 168-179, 2022,  
<http://dx.doi.org/10.18553/jmcp.2021.21290>,
- [9] Lin, Y.-J. and Gau, S.S.-F.: *Developmental Changes of Neuropsychological Functioning in Individuals with and without Childhood ADHD from Early Adolescence to Young Adulthood: A 7-Year Follow-up Study*.  
Psychological Medicine **49**(6), 940-951, 2019,  
<http://dx.doi.org/10.1017/S0033291718001599>,
- [10] Anderson, A., et al.: *Non-Negative Matrix Factorization of Multimodal MRI, FMRI and Phenotypic Data Reveals Differential Changes in Default Mode Subnetworks in ADHD*.  
Neuroimage **102**(Part 1), 207-219, 2014,  
<http://dx.doi.org/10.1016/j.neuroimage.2013.12.015>,
- [11] Sidhu, G.S.; Asgarian, N.; Greiner, R. and Brown, M.R.G.: *Kernel Principal Component Analysis for Dimensionality Reduction in FMRI-Based Diagnosis of ADHD*.  
Frontiers in Systems Neuroscience **6**, No. 74, 2012,  
<http://dx.doi.org/10.3389/fnsys.2012.00074>,
- [12] Liu, L.; Tang, S.; Wu, F.; Wang, Y.-P. and Wang, J.: *An Ensemble Hybrid Feature Selection Method for Neuropsychiatric Disorder Classification*.  
IEEE/ACM Transaction on Computational Biology and Bioinformatics **19**(3), 1459-1471, 2021,  
<http://dx.doi.org/10.1109/TCBB.2021.3053181>,
- [13] Bellec, P., et al.: *The Neuro Bureau ADHD-200 Preprocessed Repository*.  
Neuroimage **144**(Part B), 275-286, 2017,  
<http://dx.doi.org/10.1016/j.neuroimage.2016.06.034>,
- [14] Guyon, I.; Weston, J.; Barnhill, S. and Vapnik, V.: *Gene Selection for Cancer Classification Using Support Vector Machines*.  
Machine Learning **46**, 389-422, 2002,  
<http://dx.doi.org/10.1023/A:1012487302797>,
- [15] Paul, A.K.; Shill, P.C.; Rabin, M.R.I. and Murase, K.: *Adaptive Weighted Fuzzy Rule-Based System for the Risk Level Assessment of Heart Disease*.  
Applied Intelligence **48**, 1739-1756, 2018,  
<http://dx.doi.org/10.1007/s10489-017-1037-6>,

- [16] Son, L.H.; Van Viet, P. and Van Hai, P.: *Picture Inference System: A New Fuzzy Inference System on Picture Fuzzy Set*. Applied Intelligence **46**(3), 652-669, 2017, <http://dx.doi.org/10.1007/s10489-016-0856-1>,
- [17] Chen, Y.-S.; Cheng, C.-H.; Chiu, C.-L. and Huang, S.-T.: *A Study of ANFIS-Based Multi-Factor Time Series Models for Forecasting Stock Index*. Applied Intelligence **45**(2), 277-292, 2016, <http://dx.doi.org/10.1007/s10489-016-0760-8>,
- [18] Jang, J.-S.R.: *ANFIS: Adaptive-Network-Based Fuzzy Inference System*. IEEE Transactions on Systems, Man, and Cybernetics **23**(3), 665-685, 1993, <http://dx.doi.org/10.1109/21.256541>,
- [19] Fazel Zarandi, M.H.; Soltanzadeh, S.; Mohammadi, A. and Castillo, O.: *Designing a General Type-2 Fuzzy Expert System for Diagnosis of Depression*. Applied Soft Computing **80**, 329-341, 2019, <http://dx.doi.org/10.1016/j.asoc.2019.03.027>,
- [20] Castellazzi, G., et al.: *A Machine Learning Approach for the Differential Diagnosis of Alzheimer and Vascular Dementia Fed by MRI Selected Features*. Frontiers in Neuroinformatics **14**, No. 25, 2020, <http://dx.doi.org/10.3389/fninf.2020.00025>,
- [21] Shaveta A.; Vadher, R. and Chugh, B.: *A Decision-Making System for Corona Prognosis Using Fuzzy Inference System*. Journal of Fuzzy Extension and Applications **2**(4), 344-354, 2021, <http://dx.doi.org/10.22105/jfea.2021.296652.1158>,
- [22] Mirjalili, S.; Mirjalili, S.M. and Lewis, A.: *Grey Wolf Optimiser*. Advances in Engineering Software **69**, 46-61, 2014, <http://dx.doi.org/10.1016/j.advengsoft.2013.12.007>,
- [23] Emary, E.; Zawbaa, H.M.; Grosan, C. and Hassenian, A.E.: *Feature Subset Selection Approach by Grey-Wolf Optimisation*. In: Abraham, A.; Krömer, P. and Snasel, V., eds.: *Afro-European Conference for Industrial Advancement*. Advances in Intelligent Systems and Computing **334**. Springer, Cham, pp.1-13, 2015, [http://dx.doi.org/10.1007/978-3-319-13572-4\\_1](http://dx.doi.org/10.1007/978-3-319-13572-4_1),
- [24] Purushothaman, R.; Rajagopalan, S.P. and Dhandapani, G.: *Hybridizing Grey Wolf Optimization (GWO) with Grasshopper Optimization Algorithm (GOA) for Text Feature Selection and Clustering*. Applied Soft Computing **96**, No. 106651, 2020, <http://dx.doi.org/10.1016/j.asoc.2020.106651>,
- [25] Ayaz Wani, A.: *A Review of Challenges and Solutions for using Machine Learning Approaches for Missing Data*. International Journal of Engineering Applied Sciences and Technology **9**(5), 36-50, 2024, <http://dx.doi.org/10.33564/IJEAST.2024.v09i05.005>,
- [26] Karishma and Kumar, H.: *GWO-Based Energy-Efficient Workflow Scheduling for Heterogeneous Computing Systems*. Soft Computing **29**(7), 3469-3508, 2025, <http://dx.doi.org/10.1007/s00500-025-10614-y>,
- [27] van Buuren, S.: *Flexible Imputation of Missing Data*. 2<sup>nd</sup> edition. Chapman and Hall/CRC, Boca Raton, 2018, <http://dx.doi.org/10.1201/9780429492259>,
- [28] Azur, M.J.; Stuart, E.A.; Frangakis, C. and Leaf, P.J.: *Multiple Imputation by Chained Equations: What Is It and How Does It Work?* International Journal of Methods in Psychiatric Research **20**(1), 40-49, 2011, <http://dx.doi.org/10.1002/mpr.329>,

- [29] Brown, M.R.G., et al.: *ADHD-200 Global Competition: Diagnosing ADHD Using Personal Characteristic Data Can Outperform Resting State FMRI Measurements*. *Frontiers in Systems Neuroscience* **6**, No. 69, 2012, <http://dx.doi.org/10.3389/fnsys.2012.00069>,
- [30] Ghiassian, S.; Greiner, R.; Jin, P. and Brown, M.R.G.: *Using Functional or Structural Magnetic Resonance Images and Personal Characteristic Data to Identify ADHD and Autism*. *PLoS One* **11**, No. e0166934, 2016, <http://dx.doi.org/10.1371/journal.pone.0166934>,
- [31] Tan, L.; Guo, X.; Ren, S.; Epstein, J.N. and Lu, L.J.: *A Computational Model for the Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Based on Functional Brain Volume*. *Frontiers in Computational Neuroscience* **11**, No. 75, 2017, <http://dx.doi.org/10.3389/fncom.2017.00075>,
- [32] Mao, Z., et al.: *Spatio-Temporal Deep Learning Method for ADHD FMRI Classification*. *Information Sciences* **499**, 1-11, 2019, <http://dx.doi.org/10.1016/j.ins.2019.05.043>,
- [33] Riaz, A.; Asad, M.; Alonso, E. and Slabaugh, G.: *Fusion of FMRI and Non-Imaging Data for ADHD Classification*. *Computerized Medical Imaging and Graphics* **65**, 115-128, 2018, <http://dx.doi.org/10.1016/j.compmedimag.2017.10.002>,
- [34] Chen, M., et al.: *A Multichannel Deep Neural Network Model Analyzing Multiscale Functional Brain Connectome Data for Attention Deficit Hyperactivity Disorder Detection*. *Radiology: Artificial Intelligence* **2**(1), No. e190012, 2020, <http://dx.doi.org/10.1148/ryai.2019190012>,
- [35] Qiang, N., et al.: *A Novel ADHD Classification Method Based on Resting State Temporal Templates (RSTT) Using Spatiotemporal Attention Auto-Encoder*. *Neural Computing and Applications* **34**(10), 7815-7833, 2022, <http://dx.doi.org/10.1007/s00521-021-06868-w>,
- [36] Qin, Y.; Lou, Y.; Huang, Y.; Chen, R. and Yue, W.: *An Ensemble Deep Learning Approach Combining Phenotypic Data and FMRI for ADHD Diagnosis*. *Journal of Signal Processing Systems* **94**(11), 1269-1281, 2022, <http://dx.doi.org/10.1007/s11265-022-01812-0>,
- [37] Saurabh, S. and Gupta, P.K.: *Deep Learning-Based Modified Bidirectional LSTM Network for Classification of ADHD Disorder*. *Arabian Journal for Science and Engineering* **49**(3), 3009-3026, 2023, <http://dx.doi.org/10.1007/s13369-023-07786-w>,
- [38] Uyulan, C., et al.: *A Class Activation Map-Based Interpretable Transfer Learning Model for Automated Detection of ADHD from FMRI Data*. *Clinical EEG and Neuroscience* **54**(2), 151-159, 2023, <http://dx.doi.org/10.1177/15500594221122699>,
- [39] Gibson, S.J., et al.: *Using Multivariate Imputation by Chained Equations to Predict Redshifts of Active Galactic Nuclei*. *Frontiers in Astronomy and Space Sciences* **9**, No. 836215, 2022, <http://dx.doi.org/10.3389/fspas.2022.836215>,
- [40] Austin, P.C.; White, I.R.; Lee, D.S. and van Buuren, S.: *Missing Data in Clinical Research: A Tutorial on Multiple Imputation*. *Canadian Journal of Cardiology* **37**(9), 1322-1331, 2021, <http://dx.doi.org/10.1016/j.cjca.2020.11.010>,
- [41] Blazek, K.; van Zwieten, A.; Saglimbene, V. and Teixeira-Pinto, A.: *A Practical Guide to Multiple Imputation of Missing Data in Nephrology*. *Kidney International* **99**(1), 68-74, 2021, <http://dx.doi.org/10.1016/j.kint.2020.07.035>,

- [42] Mera-Gaona, M.; Neumann, U.; Vargas-Canas, R. and López, D.M.: *Evaluating the Impact of Multivariate Imputation by MICE in Feature Selection*. PLoS One **16**, No. e0254720, 2021, <http://dx.doi.org/10.1371/journal.pone.0254720>,
- [43] Takagi, T. and Sugeno, M.: *Fuzzy Identification of Systems and Its Applications to Modelling and Control*. IEEE Transactions on Systems, Man, and Cybernetics **SMC-15**(1), 116-132, 1985, <http://dx.doi.org/10.1109/TSMC.1985.6313399>,
- [44] de Campos Souza, P.V.: *Fuzzy Neural Networks and Neuro-Fuzzy Networks: A Review of the Main Techniques and Applications Used in the Literature*. Applied Soft Computing **92**, No. 106275, 2020, <http://dx.doi.org/10.1016/j.asoc.2020.106275>,
- [45] Al-ali, A.; Elharrouss, O.; Qidwai, U. and Al-Maaddeed, S.: *ANFIS-Net for Automatic Detection of COVID-19*. Scientific Reports **11**, No. 17318, 2021, <http://dx.doi.org/10.1038/s41598-021-96601-3>,
- [46] -: *Heart Disease Dataset*. <https://www.kaggle.com/johnsmith88/heart-disease-dataset>,
- [47] -: *Kidney Disease Dataset*. <https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease>,
- [48] -: *Covid-19 Dataset*. <https://github.com/burakalakuss/COVID-19-Clinical>,
- [49] -: *Breast Cancer Wisconsin Dataset*. <https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>.