A multi objective approach to evolving artificial neural networks for coronary heart disease classification

Alex Shenfield
Department of Engineering and Mathematics
Sheffield Hallam University
Sheffield
UK
Email: a.shenfield@shu.ac.uk

Shahin Rostami
Department of Computing and Informatics
Bournemouth University
Dorset
UK
Email: srostami@bournemouth.ac.uk

***Abstract*—The optimisation of the accuracy of classifiers in pattern recognition is a complex problem that is often poorly understood. Whilst numerous techniques exist for the optimisation of weights in artificial neural networks (e.g. the Widrow-Hoff least mean squares algorithm and back propagation techniques), there do not exist any hard and fast rules for choosing the structure of an artificial neural network - in particular for choosing both the number of the hidden layers used in the network and the size (in terms of number of neurons) of those hidden layers. However, this internal structure is one of the key factors in determining the accuracy of the classification.**

This paper proposes taking a multi-objective approach to the evolutionary design of artificial neural networks using a powerful optimiser based around the state-of-the-art MOEA/D-DRA algorithm and a novel method of incorporating decision maker preferences. In contrast to previous approaches, the novel approach outlined in this paper allows the intuitive consideration of trade-offs between classification objectives that are frequently present in complex classification problems but are often ignored. The effectiveness of the proposed multi-objective approach to evolving artificial neural networks is then shown on a real-world medical classification problem frequently used to benchmark classification methods.

I. Introduction

The performance of classification techniques on complex real-world problems is often reduced to a single performance metric - that of classification accuracy. However the real performance of classifiers for use in pattern recognition tasks (in terms of accuracy and efficiency) is a complex problem that is often poorly understood [10]. Whilst numerous gradient based search techniques exist for the optimisation of weights and biases in artificial neural networks (ANNs), such as the Widrow-Hoff least mean squares algorithm and Levenberg-Marquardt back propagation techniques, the literature contains little in the way of hard and fast rules for choosing the structure of an artificial neural network. Instead designers have to rely on rules of thumb for choosing both the number of hidden layers in an artificial neural network and the size (in terms of number of neurons) of those layers - factors that have been shown to have a great impact on the accuracy of a classifier [18], [27]. In recent years there has been some interest in using soft computing techniques such as evolutionary algorithms to provide a solution to this problem [33], focusing on evolving the structure of an artificial neural network to solve function approximation problems. However, complex classification problems often involve trade-offs between classification objectives that are not well suited to this kind of single objective approach.

One potential approach to satisfying trade-offs between classification objectives is to use evolutionary multi-objective optimisation (EMO) algorithms to address each of the conflicting objectives simultaneously. Typically, these EMO algorithms are run non-interactively, with a decision-maker (DM) setting the initial parameters of the algorithm and then analysing the results at the end of the execution process (which can often take hours or days to complete). This approach has been common since the late 1990s and leads to a set of potential solutions distributed across the whole trade-off surface. Whilst this can be appropriate for problems with a small number of objectives, when problems involve the consideration of many objectives (used here to refer to problems with four or more objectives) this trade-off surface can be very large. In these cases, the DM is usually more interested in a sub-region of this solution space that satisfies some domain specific criteria. However, this can be complicated by a lack of *a priori* knowledge about what trade-offs are achievable. To overcome these problems, progressive preference articulation methods have been proposed that take into account decision maker preferences [13], but these can be difficult to integrate with current state-of-the-art EMO algorithms.

The purpose of this paper is to introduce a novel evolutionary multi-objective approach to optimising the topology, weights and biases of an artificial neural network. This approach not only considers the classification accuracy, but also the potential trade-offs between classification objectives (information that is frequently disregarded when designing classification systems).

The paper is organised as follows: section II will provide a brief introduction to artificial neural networks, EMO algorithms and decision support in optimisation, and then section III will introduce the novel Weighted Z-score preference artic-

978-1-4799-6926-5/15/$31.00

ulation method and outline its integration into a state-of-the-art multi-objective evolutionary algorithm (MOEA/D-DRA). Section IV will outline the artificial neural network design problem considered in this paper, how this ANN is applied to the detection of heart disease, and how the classification can be improved by using the proposed evolutionary multi-objective approach to artificial neural network optimisation. Finally, section V will present some conclusions and outline some ideas for further work.

II. Background

A. Multi-objective optimisation using evolutionary algorithms

Many real-world optimisation problems involve the satisfaction of multiple objectives which, in a general form, can be described by a vector of objective functions f and a corresponding set of design variables x, shown below in Equation 1.

$$\min_{f}(x) = (f_1(x), f_2(x), \ldots, f_n(x)) \tag{1}$$

In real-world problems, conflicts between objectives mean that it is unlikely that a single ideal solution will be possible. Instead, the solution of a multi-objective optimisation problem often consists of a set of Pareto optimal points - where any improvement in one objective function will result in the degradation of one or more of the other objective functions. The quality of this *approximation set* can be characterised by considering three measures: proximity, diversity and pertinency [28], shown graphically in Fig. 1.

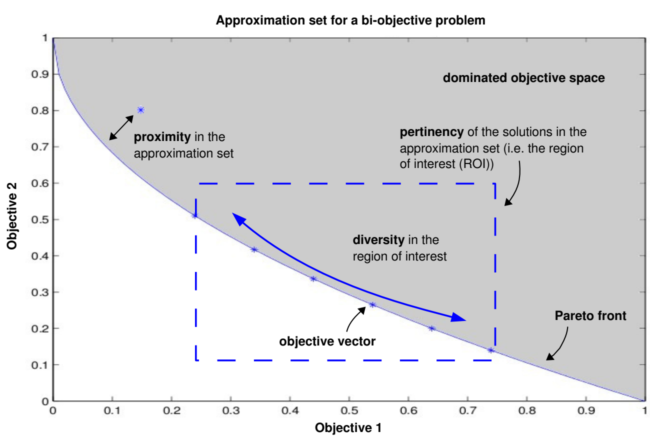

Fig. 1. Measures of approximation set quality

Conventional multi-objective optimisation techniques frequently fail to satisfy all these criteria, with methods such as goal-attainment [14] and weighted-sum [20] procedures unable to provide a diverse set of solutions to the optimisation problem. In contrast, Evolutionary Algorithms (EAs) utilise principles from natural selection to iteratively evolve a population of candidate solutions to a given problem [15] and are thus capable of presenting a diverse approximation set to a decision maker [5]. Other advantages of EAs include their robustness to multi-modal search landscapes and their use of direct objective function pay-off information in calculating the quality of candidate solutions. In addition, the population-based nature of EAs has been shown to ensure they are resilient when faced with noisy search landscapes, as each generation contains more information about the shape of the fitness landscape than would be available to conventional, non-population based methods [24].

Much of the theoretical evolutionary multi-objective optimisation literature focuses on solving problems with a small number of objectives (typically 2 and 3 objective problems). However, complex problems in the real-world frequently require the consideration of a larger number of objectives and this has led to recent interest in *many-objective*[1] *optimisation*. In a problem with many conflicting objectives, the global trade-off surface may contain many solutions that are technically Pareto-optimal but are not of interest to a decision maker [28]. An ideal many-objective optimisation procedure must therefore have the ability to filter out these irrelevant solutions.

B. Preference articulation and decision making

The role of the decision maker in evolutionary many objective optimisation is usually to choose a single compromise solution from the approximation set presented to them. Although there may be a potentially infinite number of Pareto-optimal solutions in the global trade-off surface, in practice the decision maker will usually only be interested in a small subset of these. Therefore, allowing the decision maker to focus the optimisation process on relevant areas of the search space both increases the efficiency of the search effort and reduces the amount of irrelevant information the decision maker has to consider.

The preferences of a decision maker can be incorporated into the optimisation process in three ways:

- *A posteriori*
- *A priori*
- Progressively

A posteriori methods of preference articulation involve the decision maker selecting a compromise solution from the global set of Pareto-optimal solutions found at the end of the optimisation process, whilst *a priori* and progressive preference articulation methods aim to achieve a good representation of the trade-off surface in the region of interest of the decision maker. The key advantage of *a priori* and progressive preference articulation methods is the reduction in the size of the search space explored by the optimiser because the search is focused on a sub-set of the global trade-off surface.

In *a priori* articulation of preferences the decision maker expresses their preferences before the start of the optimisation process. However, often the decision maker may not be sure of their preferences prior to optimisation and, by stating their preferences *a priori*, the decision maker may not investigate some areas of the search space that deserve attention. A better

[1]The phrase *many-objective* has been used by the operations research community to refer to problems with four or more objectives.

method is often progressive articulation of preferences, where the decision maker can alter their preferences during the search and thus incorporate information that only becomes available during the search process (such as the exact nature of trade-offs between objectives).

One of the first schemes for progressive preference articulation in EMO algorithms was introduced by [13]. It extended the Pareto-based ranking scheme used in the Multiple Objective Genetic Algorithm (MOGA) [12] to allow preferences to be expressed throughout the run of a multi-objective evolutionary algorithm. These preferences were then used in a modified version of dominance which combines the concept of Pareto-optimality with a preference operator to rank the candidate solutions according to both preference information and Pareto-dominance. This progressive preference articulation method has been used in a wide variety of engineering applications such as the optimisation of robust control strategies for gasifier power plants [17] and the design of lateral stability controllers for aircraft [32].

C. Artificial neural networks for solving classification problems

Artificial Neural Networks are a class of statistical learning algorithms inspired by the behaviour of biological neurons located in the brain and central nervous system [23], [29]. ANNs make use of a set of self-adaptive input weights and biases that are tuned by some learning algorithm to capture highly complex and non-linear underlying models of the data they are applied to. This self-adaptive nature means that they can detect complex relationships between both dependent and independent variables without prior knowledge [34].

ANNs have been widely used in a variety of pattern recognition and classification tasks. In contrast to traditional classification techniques, such as discriminant analysis, which require a good understanding of the underlying statistical model of the system that produced the data, ANNs are a "black-box" technique capable of adapting to this underlying model [37]. This makes them particularly useful in fields such as decision support for medical diagnosis [21] where their ability to adapt to the data, especially in high dimensional datasets, overcomes many of the difficulties in model building associated with conventional classification techniques such as decision trees and k-nearest neighbour algorithms [7].

A key drawback in the use of ANNs is the difficulty both in selecting appropriate network structures and in tuning the weights and biases within the network - both of which have been shown to have a large impact on the overall accuracy of classifiers [18]. Often weights and biases in ANNs are tuned using gradient descent based back propagation methods - however, these can be prone to premature convergence to local optima [16]. To overcome some of these problems there has been much interest over the last decade into evolutionary artificial neural networks (EANNs) [35]. EANNs can be configured for parametric learning (evolving the weights and biases within the ANN) or structural learning (evolving both the number of hidden layers and the number of neurons within each layer in an ANN) [36], [2]. More recently, EANNs that perform both parametric and structural learning, such as the NeuroEvolution of Augmenting Topologies (NEAT) algorithm [33], have been used for solving function approximation problems by incrementally growing and pruning the structure of an ANN. The main limitation of this kind of EANN approach is the focus on optimising a single performance metric (usually overall classifier accuracy) and thus disregarding potential trade-offs between classification objectives.

Whilst some studies into the use of multi-objective evolutionary optimisation methods in ANNs exist, they predominantly look at trade-offs between the classification accuracy of the ANN and the complexity of the network [1] rather than treating trade-offs in classification objectives separately. Although not technically a EANN method, Everson and Fieldsend [9] have used a multi-objective optimisation approach based around the Pareto archived evolution strategy (PAES) algorithm [22] to generalise receiver operator characteristic (ROC) curves to multi-class classification problems. This method is used to analyse and compare the ROC surfaces of classifiers with multiple classification objectives. However, they note that the dimensionality of this comparison increases rapidly with the number of objectives considered (for example, a classification problem with 3 target classes will require consideration of 6 dimensions). Other research into multi-objective optimisation has shown that, as the dimensionality of a multi-objective optimisation problem increases, the effectiveness of Pareto-ranking based optimisation methods decreases [19].

III. Novel preference articulation methods in advanced EMO algorithms

A. MOEA/D: decomposition based evolutionary multi-objective optimisation

The MOEA/D-DRA algorithm [39] is a state-of-the-art evolutionary multi-objective optimisation algorithm that has been shown to perform well in problems with complex Pareto fronts (such as those used in the CEC2009 test suite [40]). The approach of the basic MOEA/D algorithm [38] is to decompose a multi-objective optimisation problem into a number of single-objective optimisation subproblems using ideas taken from the mathematical programming community. These single-objective optimisation problems can then be optimised simultaneously using a population based approach with a neighbourhood information sharing model.

This state-of-the-art EMO algorithm has been integrated with a novel, two-phase preference articulation operator using weighted z-scores (described in the next section). The resulting WZ-MOEA/D-DRA algorithm is described in section III-C.

B. Weighted Z-score preference articulation

Weighted Z-score (WZ) preference articulation is a novel method of preference articulation based around the use of z-scores (or standard scores) from statistics [30], [31]. Traditionally, z-score calculations are performed by subtracting the population mean from a datum and then dividing the result by the population standard deviation as can be seen in Equation 2. Calculating the z-score in statistics requires knowing the population parameters and not just the parameters of a sample, which is often seen as unrealistic in typical statistics; however this is not an issue in EMO as it is possible to have a complete representation of the population at each generation.

$$z = \frac{(x - \mu)}{\sigma} \tag{2}$$

For the z-score to be useful for preference articulation, some modifications are made to the way z is calculated. Instead of using the population mean and population standard deviation to calculate z, the preference information that has been expressed by the DM is used (as can be seen in Equation 3) where ρ_m is the goal for a corresponding objective value x_{mn}, and N is the number of solutions in the population.

$$z_{mn} = \frac{(x_{mn} - \rho_m)}{\sqrt{\frac{\sum_{n=1}^{N}(x_{mn} - \rho_m)^2}{\mathrm{N}}}} \tag{3}$$

This will enable the calculation of z_{mn} for the objective values of each candidate solution in an approximation set, resolving the number of standard deviations each solution is from the DM's expressed region of interest (ROI), which will be a positive value when it is outside the ROI, and negative when within the ROI. Once z_{mn} is calculated for every objective value of a solution, the z_{mn} values are aggregated into a single fitness value using Equation 4.

$$V_n = \frac{\sum_{m=1}^{M} z_{mn}}{\mathrm{M}} \tag{4}$$

The mathematical procedure for the WZ preference articulation operator in its entirety is described herein. M defines the number of problem objectives whilst N defines the population size. X is an M by N matrix of entries x_{mn}, where every x_{mn} refers to a solution's objective value:

$$\mathrm{X_n} = \langle x_{1n}, x_{2n}, \ldots, x_{Mn} \rangle$$

Z is an M by N matrix of entries z_{mn}, where every z_{mn} refers to the result of the z-score preference articulation operator applied to a corresponding objective value x_{mn}:

$$\mathrm{Z_n} = \langle z_{1n}, z_{2n}, \ldots, z_{Mn} \rangle$$

To calculate Z, a preference vector P of M entries must be defined, where every entry ρ_m refers to the goal which the corresponding objective values x_m must satisfy:

$$\mathrm{P} = \langle \rho_1, \rho_2, \ldots, \rho_M \rangle$$

S is an M by N matrix of entries s_{mn} where every s_{mn} refers to a logical value indicating whether the corresponding objective value x_{mn} has satisfied the corresponding goal ρ_{mn} ($x_{mn} \leq \rho_m$):

$$S_n = \langle s_{1n}, s_{2n}, \ldots, s_{Mn} \rangle$$

where s_{mn} is calculated using:

$$s_{mn} = \begin{cases} 1, & \text{if } x_{mn} \leq \rho_m \\ 0, & \text{otherwise.} \end{cases}$$

Φ is a vector of N entries, where every ϕ_n refers to a logical value indicating whether all entries of P have been satisfied by a solution X_n.

$$\Phi = \langle \phi_1, \phi_2, \ldots, \phi_N \rangle$$

where ϕ_n is calculated by the product of the entries of S_n:

$$\phi_n = \prod_{m=1}^{M} s_{mn}$$

The scalar Ψ refers to the number of solutions X_n in the population which have satisfied the preference vector P:

$$\Psi = \sum_{n=1}^{N} \phi_n$$

T defines the required number of solutions which satisfy the preference vector before the search changes phase. Whilst $\Psi < T$ the W-phase of the WZ preference articulation operator takes effect. In this phase, the weighting $(1 - \frac{1}{M})$ is only applied to the z_{mn} value if m corresponds to the entry of Ω with the lowest value. ω_m refers to the number of solutions in the population that have satisfied the corresponding ρ_m:

$$\Omega = \langle \omega_1, \omega_2, \ldots, \omega_M \rangle$$

ω_m is the sum of columns M in the matrix S and is calculated using:

$$\omega_m = \sum_{n=1}^{N} s_{mn}$$

With the entries of Ω calculated, the M by N matrix of weighted scores E can be defined as:

$$\mathrm{E_n} = \langle \epsilon_{1n}, \epsilon_{2n}, \ldots, \epsilon_{mN} \rangle$$

where the corresponding weighted score ϵ_{mn} for each objective value x_{mn} can be calculated using:

$$\epsilon_{mn} = \begin{cases} z_{mn}\left(1 - \frac{1}{\mathrm{M}}\right) & \text{if } f(\omega_m, S_{mn}) = 0 \\ z_{mn} & \text{otherwise.} \end{cases}$$

where z_{mn} and ω_m are first normalised to real values between 0 and 1:

$$z_{mn} = f(|z_{mn}|, |Z_m|)$$

using the function $f(k, K)$ where:

$$f(k, K) = \frac{k - min(K)}{max(K - min(K))}$$

The initial calculation of z_{mn} is the same in both phases (W-phase and Z-phase) and is defined in Equation 3. The final score W_n of a single solution is the aggregation of the corresponding ϵ_{mn} entries:

$$W_n = \frac{\sum_{m=1}^{M} \epsilon_{mn}}{\mathrm{M}} \tag{5}$$

This two-phase method attempts to move the search towards the production of solutions that are close in proximity to the ROI and within it, but does not attempt to minimise the solutions beyond the edges of the ROI. When the number of solutions within the ROI has satisfied the threshold ($\Psi \geq T$) the Z-phase takes effect. This phase uses Equation 3 to calculate Z_n and then Equation 4 to aggregate the scores into the scalar V_n, this is because there are adequate solutions (defined by T) that have satisfied all entries of P. These solutions can then be further minimised within the ROI.

C. WZ-MOEA/D-DRA

The weighted z-score preference articulation operator described in the previous section has been incorporated into the state-of-the-art MOEA/D-DRA algorithm [39], in order to allow selection pressure towards a desired ROI during the optimisation process. The new preference driven algorithm (WZ-MOEA/D-DRA) has been benchmarked on a selection of synthetic test problems and applied successfully to a real-world many-objective problem regarding the optimisation of classifiers for concealed weapon detection [30]. WZ-MOEA/D-DRA has been shown to offer robust performance on complex many-objective problems consisting of less than seven objectives.

WZ-MOEA/D-DRA operates in one of two phases (W-phase and Z-phase) dictated by the WZ preference articulation operator, which take effect depending on when certain criteria are satisfied, allowing the optimisation process to efficiently spend the function evaluation budget depending on the current optimisation context.

Whilst the number of solutions satisfying the preference vector P is below the threshold ($\Psi < T$) the W-phase of the WZ preference articulation operator takes effect. In this phase the MOEA/D-DRA's utility selection is replaced with a selection of solutions based on their W_n score calculated using Equation 5.

If during the optimisation process the threshold ($\Psi \geq T$) is satisfied then the Z-phase of the WZ preference articulation operator takes effect, whilst in this phase a modified implementation of MOEA/D-DRA's utility selection is used, where the edging sub-problems are no longer considered as elite and solutions that do not satisfy ($\phi_n = 0$) the DM's expressed preferences P are discarded.

Using these two phases WZ-MOEA/D-DRA is able to get close in proximity to the DM's expressed ROI within a small number of function evaluations, and then produce solutions within the ROI and minimise solutions whilst retaining the diversity features of MOEA/D-DRA.

The contributing hypervolume indicator [8] is used post-optimisation in order to cull the approximation set to a more digestible size, in order to allow the DM to make a decision without being overwhelmed with choice. This process has been illustrated in Fig. 2.

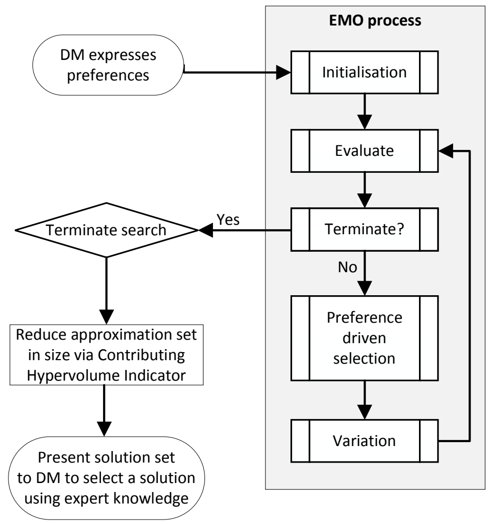

Fig. 2. The execution life-cycle of a EMO process allowing for the incorporation of decision maker preferences

IV. Classifying the severity of instances of heart disease

A. Problem description

Coronary Heart Disease (CHD) is one of the leading causes of death both in the UK and globally [26]. It is responsible for approximately 73,000 deaths per in the UK every year [3] and, in the UK alone, it is estimated that 2.5 million people are living with the condition. To accurately diagnose the presence and severity of coronary heart disease generally involves the use of a coronary angiogram - an expensive and invasive procedure that is unsuitable for large scale screening of the population. One possible solution to this is to use computational methods of predicting heart disease instances to provide an initial estimate of the likely-hood of CHD.

Detrano et. al. [6] collected heart disease data from 303 cases at V.A. Medical Center, Long Beach & Cleveland Clinic Foundation to build a discriminant function model for estimating probabilities of coronary heart disease. This data set is widely used in the classification literature to benchmark new classifiers [4], [25] and consists of 76 problem attributes in total. The majority of studies based on this dataset consider 14 of these attributes, summarised in [25]. In this paper a subset of 297 cases from this dataset is considered (discarding the 6 cases where the attribute information is incomplete).

This data set categorises the severity of heart disease from 0 (no heart disease) to 1 through 4 (increasing severity of heart disease). Although this data set has been widely used in the classification literature, all published experiments have focused on distinguishing the presence of heart disease (1-4) from the absence (0). In contrast, our multi-objective approach to evolving artificial neural networks for classification tasks aims to not only classify presence / absence of heart disease, but also to identify the severity of heart disease and minimise the number of mis-classifications.

The 6 objectives used in our approach are shown in Table I. Note that these have been converted into minimisation objectives for the purpose of optimisation.

B. Encoding the problem

In order to use evolutionary methods to optimise the topology and weights of the ANN classifier for heart disease

TABLE I
PERFORMANCE OBJECTIVES

Objective 1	Classified normal correctly
Objective 2	Misclassified heart disease as normal
Objective 3	Classified heart disease correctly
Objective 4	Misclassified normal as heart disease
Objective 5	Classified mild heart disease correctly
Objective 6	Classified severe heart disease correctly

detection, the ANNs topology and weights must be encoded into a real-valued chromosome, which can then be subjected to the various evolutionary operators used in the optimisation process and then decoded for evaluation. Fig. 3 illustrates the chromosome structure used to store the encoding of an ANN with 5 output neurons, a maximum of 3 hidden layers, and a minimum of 13 input neurons.

Parameter boundaries are also required to restrict the number of hidden layers, neurons per hidden layer, and ranges for the weights and biases within a lower and upper limit. All hidden layers but the last can contain a number of neurons ranging from none to twice the number of input neurons, as seen in Equation 6, and the last hidden layer must contain a minimum of neurons equal to the number of input neurons as seen in Equation 7. This means each candidate network generated by the optimiser must have at least one hidden layer, preventing the generation of benign networks which would waste function evaluations throughout the entire optimisation process. Finally, each weight and bias is restricted to the same boundary shown in Equation 8.

$$b(1...(HL-1)) = \{x \in \mathbb{Z} \mid 0 \leq x \leq 2i\} \tag{6}$$

$$b(HL) = \{x \in \mathbb{Z} \mid i \leq x \leq 2i\} \tag{7}$$

$$w = \{x \in \mathbb{R} \mid -5 \leq x \leq 5\} \tag{8}$$

For the ANN used in this network, each candidate solution contains 1906 variables, with the first 3 defining the number of hidden layers and the number of neurons on each respectively, the following 338 variables defining the weights for the input layer, 676 for the first and second hidden layer, and 130 for the third and final hidden layer.

Regardless of the topology of the candidate solution ANN (which in this case is defined by the first three genes of the encoded chromosome) the maximum number of weights and biases will be stored with each chromosome; however, not all genotypes will manifest themselves and be expressed as phenotypes as only the weights and biases required to configure the candidate solutions ANN topology will be decoded and used. These unused weights and biases will remain unexpressed in the phenotype until the first three genes allow them to manifest and can go through many generations as dormant genes. This introduces the interesting feature of atavism[2].

At each function evaluation, a chromosome is decoded from its encoded state (as described in Fig. 3) and used to instantiate an ANN. This ANN is then used to classify the training data and the results of this assessed against the performance objectives specified in Table I. Following the completion of the optimisation process, the final generation of candidate solutions is decoded and used to create ANNs which are then run on the unseen testing data to obtain the final results (i.e. those shown in Figs. 4, 5, and 6).

C. *Optimisation results and discussion*

The ANN encoding was optimised using the WZ-MOEA/D-DRA algorithm described in Section III with the parameters shown in Table II. The performance was then evaluated using the objectives specified in Table I. A real-valued representation for the ANN encoding parameters shown in Fig. 3 was used, since Fogel and Ghozeil [11] have shown that there is no intrinsic advantage in choosing one bijective representation over another, although particular representations may be more computationally tractable or efficient for certain problems. As a consequence of this, modern EMO practice emphasises choosing a representation that is appropriate for the problem under consideration [24] and, in this application, the ANN parameters in our encoding are real-valued.

TABLE II
ALGORITHM PARAMETERS

Population size	50
Maximum generations	250
Z-score threshold (T)	5
MOEA/D neighbourhood	30

Fig. 4 shows that there is a clear trade-off between maximising the accuracy of positively diagnosing heart disease and minimising the misdiagnosis of cases of heart disease, as well as between maximising the accuracy of diagnosing the absence of heart disease and misdiagnosing the absence of heart disease. These results were taken from 25 runs of the optimiser with no goals specified. Each data point on Fig. 4 represents a single candidate solution ANN from the final generation produced by the optimiser. This set of ANNs was then run on an unseen testing data set to produce the points in Fig. 4.

Fig. 5 shows the initial parallel coordinate plot presented to a decision maker representing the potential trade-offs between the classification objectives (as described in Table I). This was generated using the goals shown in Table III. The highlighted solutions are the 6 best solutions from this set in terms of the contributing hyper-volume metric.

Having seen what is achievable from Fig. 5, it is possible to tighten some of the goals (using domain specific knowledge)

[2] In biology, atavism is a tendency for evolutionary traits to lie dormant (for example, remaining present in DNA but not being expressed as a phenotypical feature) but remain intact. In these cases it is possible for a fault in the genetic feature suppressing the trait (possibly through a mutation of that gene) to lead to it reasserting itself.

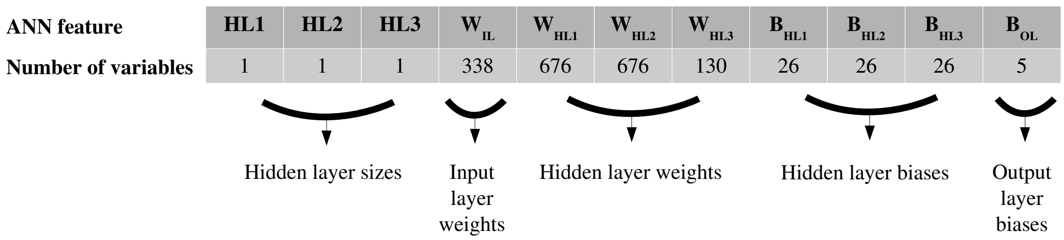

Fig. 3. Encoded chromosome for the six-objective ANN consisting of 3 hidden layers (HL), an input layer (IL), 5 neurons on the output layer (OL), and associated biases, totalling to 1906 variables

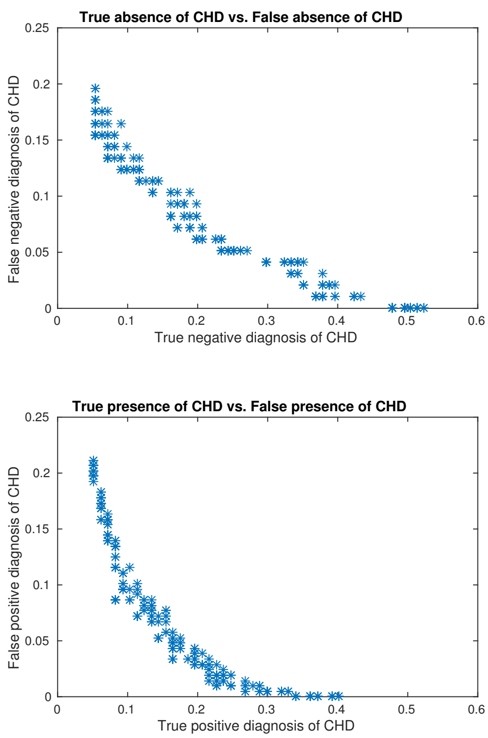

Fig. 4. Trade-offs between classification objectives

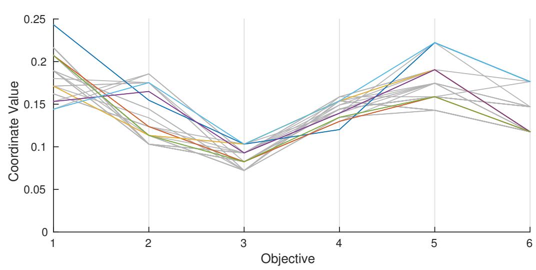

Fig. 5. Parallel coordinate plot of results

TABLE III
INITIAL GOALS

Objective	1	2	3	4	5	6
Goal (accuracy in %)	>50	<20	>50	<30	>60	>40

to reduce the number of solutions presented to the decision maker. In this case, it is better to err on the side of caution so a decision maker would be prepared to accept an increase in the percentage of false positive diagnoses of heart disease if it results in a lower percentage of cases of heart disease missed. Fig. 6 shows the revised parallel coordinate plot presented to a decision maker using a stricter set of goals (shown in Table IV). The highlighted solutions are again the 6 best solutions from this set in terms of the contributing hyper-volume metric. Note that, in this figure, there are many less solutions presented to the decision maker. Table V shows a summary of the results from 100 independent runs of the optimisation routine. Over these 100 runs, the optimiser found solutions within the stricter ROI (shown in Table IV) 93 times (and within the original ROI every time), proving that the proposed evolutionary multi-objective optimisation of ANNs exhibits robust performance.

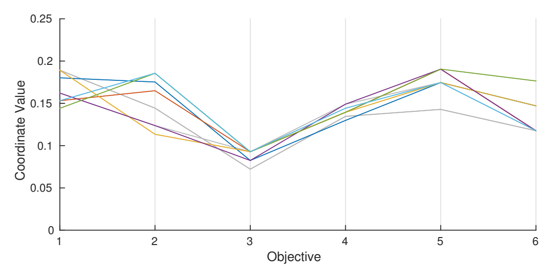

Fig. 6. Parallel coordinate plot of results with stricter goals

TABLE IV
STRICTER GOALS

Objective	1	2	3	4	5	6
Goal (accuracy in %)	>80	<20	>90	<15	>60	>40

TABLE V
SUMMARY OF 100 RUNS OF THE OPTIMISATION ALGORITHM

Average number of solutions in ROI	13.79
Standard deviation of number of solutions in ROI	11.86
Maximum number of solutions in ROI	41
Minimum number of solutions in ROI	0

V. CONCLUSIONS AND FURTHER WORK

In this paper a novel method of optimising the weights, biases and topology of an artificial neural network by considering classification trade-offs in a multi-objective way has been introduced. This multi-objective optimiser is based around the state-of-the-art MOEA/D-DRA optimisation algorithm and the recently introduced Weighted Z-score method of handling decision maker preferences.

The application of the WZ-MOEA/D-DRA optimisation algorithm to the training and optimisation of the topology, weights and biases of an ANN intended for use in the diagnosis of heart disease has been presented. It has been shown that, by handling classification tasks with multiple target classes in a multi-objective way, it is possible to not only achieve good classification accuracy overall but also minimise misclassifications. This multi-objective optimisation technique with the integration of preferences has been shown to provide the decision maker with a number of solutions (trained ANNs) with trade-offs that are well distributed across the Pareto front. The decision maker can then select an optimised solution which balances false positive diagnoses of heart disease with cases where heart disease is missed.

References

[1] H. A. Abbass. Speeding up backpropagation using multiobjective evolutionary algorithms. *Neural Computation*, 15(11):2705–2726, 2003.

[2] P. J. Angeline, G. M. Saunders, and J. B. Pollack. An evolutionary algorithm that constructs recurrent neural networks. *Neural Networks, IEEE Transactions on*, 5(1):54–65, 1994.

[3] British Heart Foundataion. *Cardiovascular Disease Statistics 2014*. British heart foundation centre on population approaches for non-communicable disease prevention, Oxford, 2014.

[4] R. Das, I. Turkoglu, and A. Sengur. Effective diagnosis of heart disease through neural networks ensembles. *Expert Systems with Applications*, 36(4):7675 – 7680, 2009.

[5] K. Deb. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons, New York, 2001.

[6] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher. International application of a new probability algorithm for the diagnosis of coronary artery disease. *The American Journal of Cardiology*, 64(5):304 – 310, 1989.

[7] S. Dreiseitl and L. Ohno-Machado. Logistic regression and artificial neural network classification models: a methodology review. *352-359*, 35(5-6), 2002.

[8] M. Emmerich, N. Beume, and B. Naujoks. An emo algorithm using the hypervolume measure as selection criterion. In *Evolutionary Multi-Criterion Optimization*, pages 62–76. Springer, 2005.

[9] R. M. Everson and J. E. Fieldsend. Multi-class roc analysis from a multi-objective optimisation perspective. *Pattern Recognition Letters*, 27(8):918–927, 2006.

[10] J. C. Fernandez Caballero, F. J. Martínez, C. Hervás, and P. A. Gutiérrez. Sensitivity versus accuracy in multiclass problems using memetic pareto evolutionary neural networks. *Neural Networks, IEEE Transactions on*, 21(5):750–770, 2010.

[11] D. B. Fogel and A. Ghoziel. A note on representations and variation operators. *IEEE Transactions on Evolutionary Computation*, 1(2):159–161, 1997.

[12] C. M. Fonseca and P. J. Fleming. Genetic algorithms for multi-objective optimization: Formulation, discussion and generalization. In S. Forrest, editor, *Proceedings of the Fifth International Conference on Genetic Algorithms*, pages 416–423. Morgan Kaufmann, 1993.

[13] C. M. Fonseca and P. J. Fleming. Multiobjective optimization and multiple constraint handling with evolutionary algorithms - Part I: A unified formulation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 28(1):26–37, 1998.

[14] F. W. Gembicki. *Vector Optimization for Control with Performance and Parameter Sensitive Indices*. PhD thesis, Case Western Reserve University, Cleveland, Ohio, 1974.

[15] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA., 1989.

[16] M. Gori and A. Tesi. On the problem of local minima in backpropagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(1):76–86, 1992.

[17] I. Griffinn, P. Schroder, A. J. Chipperfield, and P. J. Fleming. Multi-objective optimization approach to the alstom gasifier problem. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 214(6):1453–1469, 2000.

[18] M. T. Hagan, H. B. Demuth, M. H. Beale, et al. *Neural network design*. Pws Pub. Boston, 1996.

[19] E. J. Hughes. Evolutionary many-objective optimisation: many once or one many? In *Evolutionary Computation, 2005. The 2005 IEEE Congress on*, volume 1, pages 222–227. IEEE, 2005.

[20] C.-L. Hwang and A. S. M. Masud. *Multiple Objective Decision Making - Methods and Applications*, volume 164 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin, 1979.

[21] T. Iidaka. Resting state functional magnetic resonance imaging and neural network classified autism and control. *Cortex*, 63(0):1047–1057, 2015.

[22] J. D. Knowles and D. W. Corne. Approximating the nondominated front using the Pareto archived evolution strategy. *Evolutionary Computation*, 8(2):149–172, 2000.

[23] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[24] Z. Michalewicz and D. B. Fogel. *How to Solve It: Modern Heuristics*. Springer, Berlin, 2000.

[25] T. Nguyen, A. Khosravi, D. Creighton, and S. Nahavandi. Classification of healthcare data using genetic fuzzy logic system and wavelets. *Expert Systems with Applications*, 42(4):2184 – 2197, 2015.

[26] NHS. Nhs choices, 2015. Viewed 10 January 2015.

[27] J. D. Paola and R. A. Schowengerdt. The effect of neural-network structure on a multispectral land-use/land-cover classification. *Photogrammetric Engineering and Remote Sensing*, 63(5):535–544, 1997.

[28] R. C. Purshouse. *On the Evolutionary Optimisation of Many Objectives*. PhD thesis, Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK, S1 3JD, 2003.

[29] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386–408, 1958.

[30] S. Rostami. *Preference Focussed Many-Objective Evolutionary Computation*. PhD thesis, Manchester Metropolitan University, 2014.

[31] S. Rostami, D. O'Reilly, A. Shenfield, and N. Bowring. A novel preference articulation operator for the evolutionary multi-objective optimisation of classifiers in concealed weapons detection. *Information Sciences*, 295:494 – 520, 2015.

[32] A. Shenfield, P. J. Fleming, and M. Alkarouri. Computational steering of a multi-objective evolutionary algorithm for engineering design. *Engineering Applications of Artificial Intelligence*, 20(8):1047–1057, 2007.

[33] K. Stanley and R. Miikkulainen. Evolving neural networks through augmenting topologies. *Evolutionary computation*, 10(2):99–127, 2002.

[34] J. V. Tu. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, 49(11):1225–1231, 1996.

[35] X. Yao. A review of evolutionary artificial neural networks. *International journal of intelligent systems*, 8(4):539–567, 1993.

[36] X. Yao and Y. Liu. Making use of population information in evolutionary artificial neural networks. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 28(3):417–425, 1998.

[37] G. P. Zhang. Neural networks for classification: a survey. *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, 30(4):451–462, 2000.

[38] Q. Zhang and H. Li. Moea/d: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712–731, 2007.

[39] Q. Zhang, W. Liu, and H. Li. The performance of a new version of moea/d on cec09 unconstrained mop test instances. In *Proceedings of the Congress on Evolutionary Computation (CEC) 2009*, pages 203–208. IEEE, 2009.

[40] Q. Zhang, A. Zhou, S. Zhao, P. N. Suganthan, W. Liu, and S. Tiwari. Multiobjective optimization test instances for the cec 2009 special session and competition. Technical Report CES-887, School of Computer Science and Electrical Engineering, University of Essex, 2008.