Sheffield Hallam University

Multi-objective evolution of artificial neural networks in multi-class medical diagnosis problems with class imbalance

SHENFIELD, Alex <http://orcid.org/0000-0002-2931-8077> and ROSTAMI, Shahin <http://orcid.org/0000-0001-8977-1239>

Available from Sheffield Hallam University Research Archive (SHURA) at:

https://shura.shu.ac.uk/16213/

This document is the Submitted Version

Citation:

SHENFIELD, Alex and ROSTAMI, Shahin (2017). Multi-objective evolution of artificial neural networks in multi-class medical diagnosis problems with class imbalance. In: CIBCB 2017 : IEEE International Conference on Computational Intelligence in Bioinformatics and Computational Biology, Manchester, 23-25 August 2017. [Conference or Workshop Item]

Copyright and re-use policy

See http://shura.shu.ac.uk/information.html

Sheffield Hallam University Research Archive
http://shura.shu.ac.uk

Multi-objective evolution of artificial neural networks in multi-class medical diagnosis problems with class imbalance

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*—This paper proposes a novel multi-objective optimisation approach to solving both the problem of finding good structural and parametric choices in an ANN and the problem of training a classifier with a heavily skewed data set. The state-of-the-art CMA-PAES-HAGA multi-objective evolutionary algorithm [41] is used to simultaneously optimise the structure, weights, and biases of a population of ANNs with respect to not only the overall classification accuracy, but the classification accuracies of each individual target class. The effectiveness of this approach is then demonstrated on a real-world multi-class problem in medical diagnosis (classification of fetal cardiotocograms) where more than 75% of the data belongs to the majority class and the rest to two other minority classes. The optimised ANN is shown to significantly outperform a standard feed-forward ANN with respect to minority class recognition at the cost of slightly worse performance in terms of overall classification accuracy.**

I. Introduction

The performance of classification techniques on complex multi-class real-world problems is often reduced to a single performance metric such as classification accuracy. However, overall classification accuracy can be a problematic performance metric — particularly in the classification of imbalanced data sets (commonly found in machine learning problems and particularly those in the medical domain). This is because overall classification accuracy assumes that the misclassification cost for each target class is equal (whereas in reality this in unlikely to be the case). This is further complicated in multi-class classification problems where more robust performance metrics such as the area under the receiver operator characteristics curve cannot easily be applied.

Another issue in the design of classifiers is finding both the optimal architecture and parameters — something that has a significant effect on classification performance. Whilst many local search techniques exist for refining weights and biases, e.g. in an artificial neural network (ANN), they are all prone to premature convergence to local optima. There also exists very little guidance in the literature as to how to choose the structure of an ANN (in terms of both the number of hidden layers and the number of neurons in those hidden layers). In recent years there has been increasing interest in addressing this problem of finding optimal ANN topologies and parameters through evolutionary methods, but most implementations of evolutionary artificial neural networks (EANNs) focus on optimising a single performance objective rather than considering potential trade-offs between the performance of the classifier on multiple target classes.

The purpose of this paper is to introduce a novel multi-objective optimisation approach to the design of ANN classifiers based around the state-of-the-art Covariance Matrix Adaptation Pareto Archieved Evolution Strategy with Hypervolume sorted Adaptive Grid Algorithm (CMA-PAES-HAGA). This novel multi-objective approach not only addresses the problem of choosing optimal ANN topologies and parameters but, by considering trade-offs in classification performance between multiple target classes, also addresses the problems associated with class imbalance.

The paper is organised as follows: Section II provides a brief introduction to ANNs, evolutionary multi-objective optimisation (EMO) algorithms and decision support in optimisation, followed by Section III which introduces the CMA-PAES-HAGA and outline its use in the structural and parametric optimisation of ANNs. Section IV will outline the ANN design problem considered in this paper, how this ANN is applied to the detection of anomalies in fetal cardiotocograms, and how the classification can be improved by using the proposed evolutionary multi-objective approach to ANN optimisation. Finally, Section V will present some conclusions drawn from this work.

II. Background

A. Artificial neural networks

ANNs have been widely used in solving real-world classification problems in a variety of application domains. ANNs use a set of interconnected data processing nodes with associated weights and biases that are capable of being tuned by some learning algorithm to capture highly complex and nonlinear relationships between both dependent and independent variables with prior knowledge [47].

Unlike many other classification methods, ANNs are simple to apply to both multi-class and binary classification problems. Other commonly used classification techniques such

as support vector machines (SVMs) and logistic regression require the use of complex multi-class strategies such as one-vs-all or all-vs-all to extend the binary classification to the multi-class case [39]. In feed-forward multilayer perceptrons it is just a case of defining the number of output neurons equal to the number of classes that the problem considers.

One disadvantage of applying ANNs to classification problems is the difficulty in optimising the weights and biases of the network, and also selecting appropriate network topologies in the first place — both of which have substantial impact on the performance of the ANN and are highly problem dependent [25]. Gradient descent based back-propagation methods are commonly used in the training process to adjust the weights and biases of the interconnected artificial neurons so as to minimise some error function, but these can be prone to premature convergence to local optima [24]. In addition to this, selecting the appropriate ANN structure is somewhat of a "black-art", with little definitive guidance in the literature [9], [10]. The network structure must be sufficiently complex to capture the underlying relationships between variables, but not so complex that over-fitting to the training data becomes a problem. Some authors have proposed strategies for dynamically growing the ANN structure as part of the training process [6], whilst others have proposed pruning the network [38]. However, both of these approaches add substantial complexity to the training process, resulting in much slower convergence, and tend to be gradient descent based and therefore prone to being trapped in local minima [21].

Over the last two decades, there has been increasing interest in the use of EANNs which are capable of addressing some of the above shortcomings. EANNs can be configured for *structural* learning (evolving the number and size of the hidden layers in the network) or *parametric* learning (optimising the weights and biases within the network) [49]. More recently, algorithms which evolve both the weights and the structure of an ANN, such as the NeuroEvolution of Augmenting Topologies (NEAT) [45], have shown promise in solving function approximation problems. However, these approaches tend to focus on optimising a single performance metric rather than taking into account any potential trade-offs between classification objectives. Stanley and Miikulainen [45] apply NEAT to the problem of double pole balancing and use the number of time steps that the pole remains balanced as the fitness. Other authors have applied NEAT to classification tasks [12] using fitness functions based around the overall classification accuracy.

Whilst some studies into multi-objective EANNs do exist, they primarily focus on the trade-offs between classifier performance and structural complexity [1], [2] rather than considering the trade-offs between the classification accuracy of different classes separately. Caballero et al. [21] proposes a multi-objective approach to evolving ANNs for multi-class problems based around NSGA-II [15]. However, instead of considering each class as an optimisation objective, they focus on the trade-off between overall classification accuracy and sensitivity.

Everson and Fieldsend [18] have proposed an interesting multi-objective optimisation approach based around the Pareto archived evolution strategy (PAES) algorithm [32] to generalise receiver operator characteristic (ROC) curves to multi-class classification problems. Although not an EANN, this method does consider the different trade-offs between multiple classification objectives in a multi-class problem. 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 due to the need to consider both sensitivity and specificity for each class). Other research into multi-objective optimisation has shown that, as the number of objectives in a multi-objective optimisation problem increases, the effectiveness of Pareto-ranking based optimisation methods decreases because candidate solutions are more likely to be non-dominated in one or more problem objectives [30].

B. Class imbalance

It is common in many machine learning problems, particularly those in the medical domain, for there to be significant differences in the prior class probabilities (i.e. class imbalances) [29], [46], [48]. This class imbalance problem often causes significant issues for classifiers, with the probability of a sample being assigned to the majority class often being overstated [46]. Imbalanced data sets have been shown to significantly compromise the performance of most commonly used classification learning algorithms [46].

Over the last decade or so there has been much interest in addressing this class imbalance problem. Current state-of-the-art techniques primarily fall into three categories [22]:

- **Algorithmic modifications** that aim to bias the learning towards the minority class (such as those proposed in [33]).
- **Data sampling and augmentation strategies** that aim to rebalance class distributions by resampling the data set (such as those proposed in [11] and [28]).
- **Cost sensitive learning approaches** where an *a priori* set of misclassification costs is incorporated into the classifier (see [17] and [50] for examples).

Out of these three approaches, resampling the data set to rebalance the classes is the only one that doesn't require extensive problem specific knowledge that may not be readily available (such as the costs of misclassification). However, many of the resampling and data augmentation techniques that have been proven to be effective for binary classification problems (such as SMOTE [11] and ADASYN [28]) can actually impair classifier performance when applied to multi-class problems [3]. Some research into resampling techniques for multi-class problems does exist [20], [3], but the majority of experimental results only consider low dimensional data sets.

C. Multi-objective optimisation using evolutionary algorithms

Many real-world optimisation problems involve the satisfaction of multiple conflicting objectives. This general form can be described by a vector f consisting of m objective functions and a corresponding set of decision variables x, as shown in (1).

$$\min f(x) = (f_1(x), \ldots, f_m(x)) \tag{1}$$

In the case of conflicting objectives, there is unlikely to be a single ideal solution. Instead, the solution to this kind of multi-objective optimisation problem leads to a family of Pareto optimal points, where any improvement in the value of one objective function will result in the degradation of one or more of the other objective functions.

A set of non-dominated solutions[1] generated by an optimisation algorithm is known as an *approximation set* [51]. The quality of the solutions in this approximation set can be characterised by the *proximity* of solutions to the true Pareto front, the *diversity* of the solutions in terms of both the extent across the true Pareto front and the uniformity of the distribution of solutions, and the *pertinence* of the solutions to the decision maker's preferences [37]. These measures are illustrated graphically in Figure 1.

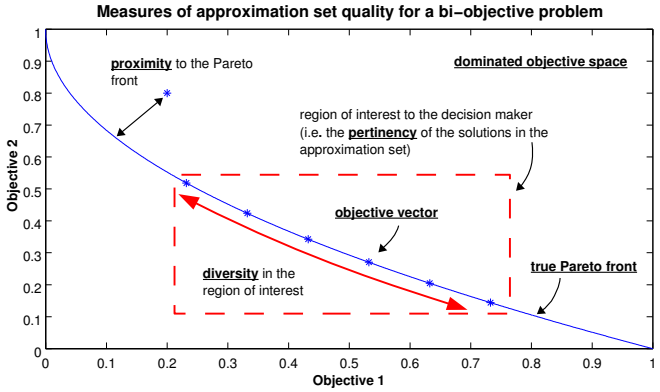

Fig. 1: A graphical representation of approximation set quality

Conventional multi-objective optimisation techniques often fail to satisfy all these quality measures. For example, the goal-attainment method [23] and the weighted-sum method [31] both only provide a single solution to the optimisation problem, failing to provide a diverse distribution of solutions across the true Pareto front. However, evolutionary algorithms (EAs) are well suited to the multi-objective optimisation of conflicting objectives because they search a population of candidate solutions [14] and are therefore capable of maintaining diversity in the approximation set throughout the optimisation process.

Other advantages of EAs include the stochastic nature of the optimisation process and the use of objective function payoff information directly (rather than derivative information or other auxiliary knowledge) which help to ensure the robustness of EAs by reducing the chances of becoming trapped in local optima. This, in conjunction with the greater amount of information about the fitness landscape contained in a population of solutions, ensures that EAs are applicable to many areas in which conventional optimisation methods struggle.

Real-world applications of optimisation, particularly in the fields of medical decision making and bioinformatics, often require a larger number of objectives to be dealt with than is typically considered in the EMO literature, leading to a growing interest amongst the research community in the area of *many-objective optimisation*[2]. The increased number of objectives in many-objective optimisation problems often mean that the trade-off surface contains a significant number of solutions that, whilst non-dominated with respect to the other solutions in the approximation set, are of no interest to a potential decision maker [37]. However, much of the time a potential decision maker may not be aware of what potential trade-offs exist until the optimisation process starts and therefore is not able to articulate any preferences *a priori*. It is therefore increasingly important that the optimiser is capable of presenting a diverse approximation set to the decision maker that captures as much of the trade-off surface as possible.

III. An Evolution Strategy for Artificial Neural Network Optimisation

The optimisation algorithm employed for this experiment is the Covariance Matrix Adaptation Pareto Archived Evolution Strategy with Hypervolume-sorted Adaptive Grid Algorithm (CMA-PAES-HAGA) [41]. This algorithm incorporates the problem space exploration of Covariance Matrix Adaption [27], and a selection mechanism based on the hypervolume indicator by means of a fast and approximated algorithm [42]. This combination allows CMA-PAES-HAGA to maintain a diverse population of solutions from iteration to iteration of the optimisation process whilst using variation operators to converge towards the theoretical optima of the problem, i.e. the true Pareto-optimal front.

The result is an algorithm which can produce an approximation set consisting of solutions, in this case trained ANNs, which offer a range of trade-offs in terms of the considered problem objectives. This allows the expert to act as a decision maker and make an informed decision as to which ANN offers suitable performance for the application, instead of using assumptions and making decisions for them.

All the details regarding CMA-PAES-HAGA are reported in [41], however for the sake of clarity the pseudo-code describing the working principles of CMA-PAES-HAGA has been presented in Algorithm 1.

The algorithm parameters used for this experiment were: number of problem objectives $M = 4$, number of problem variables $V = 989$, parent population size $\mu = 50$, and an offspring population size $\lambda = 50$ (CMA-PAES-HAGA recommends small population sizes. The smallest number of grid divisions possible in CMA-PAES-HAGA, $\delta = 2$, was selected as the population size and number of objectives are not large enough to result in infeasible computational time when calculating the hypervolume indicator. The optimisation process was restricted to a maximum of $10,000$ function evaluations, i.e. ANN evaluations.

Before using CMA-PAES-HAGA for the optimisation process, the data-set was subjected to Principal Component Analysis (PCA) to reduce the number of data attributes, and therefore the number of variables which were to be considered throughout the optimisation process. The fraction of variance

[1] I.e. those solutions where there is no other solution in the population that is superior in all objectives

[2] The phrase *many-objective* has been suggested in the operations research (OR) community to refer to problems with more than the standard two or three objectives [19].

Algorithm 1 CMA-PAES-HAGA execution cycle

// initialise the generation counter and extreme values
// vector
$g \leftarrow 0$
$\mathrm{E} \leftarrow \langle \varepsilon_1 = 0, \varepsilon_2 = 0, \ldots, \varepsilon_M = 0 \rangle$
// initialise parent population, where V contains the
// solutions in the search space and X contains the
// vectors of objective values
initialise parent population X, V
while termination criteria not met **do**
    for $n = 1, ..., \lambda$ **do**
        // variation of solutions
        $V_n^{'} \leftarrow V_n$
        $V_n^{'} \leftarrow V_n^{'} + \sigma_n \cdot \mathcal{N}(0, C_n)$
        // check solution is within bounds
        if $v_i^{(L)} \not\leq v_{in}^{'} \not\leq v_i^{(U)}$ **then**
            $v_{in}^{'} = \begin{cases} v_i^{(U)} & \text{if } v_{in}^{'} > v_i^{(U)} \\ v_i^{(L)} & \text{otherwise} \end{cases}$
        end if
    end for
    // evaluate solution and update extreme values
    $X_n^{'} \leftarrow f(V_n^{'})$
    $X^* = X \cup X^{'}$
    for $m = 1, ..., M$ **do**
        $\varepsilon_m = \begin{cases} x_{mn}^* & \text{if } x_{mn}^* > \varepsilon_m \\ \varepsilon_m & \text{otherwise} \end{cases}$
    end for
    // selection routine as in [41]
    $X, V \leftarrow$ HypervolumeSortedAGA(X^*, E)
    // variation routine as in [27]
    CMAParameterUpdate()
    $g \leftarrow g + 1$
end while

contribution was set to 0.02, which reduced the data-set from consisting of 21 attributes to only 12. One of the key benefits of this dimensionality reduction is reducing overfitting in the classifier design - though care must be taken in reducing the problem dimensions as there is a trade-off between reducing the network size (and therefore reducing chances of overfitting) vs increasing chance of removing potentially distinguishing features that just happen to have low variance.

IV. Development of a decision support system for automated analysis of cardiotocograms

A. Problem description

Complications during childbirth accounted for 256 perinatal deaths in the UK in 2014 [35] and, in more than half of these cases, improvements in care have been identified that may have made a difference to the outcome [16]. One area of criticism that has been repeatedly highlighted is the failure to correctly interpret fetal heart rate traces [40], [16]. Several studies [7], [36] have drawn attention to the poor reproducibility in the visual analysis of intrapartum cardiotocograms (CTGs), with Alfirevic et al. [4] suggesting that fetal CTG monitoring is more likely to increase the rate of caesarean delivery rather than improve clinical outcomes.

Several automated systems exist for the analysis of cardiotocograms [13], [5], [34], [8]. However, whilst these reduce the subjectiveness of the visual analysis of the CTG data, they still require expert analysis by obstetricians to interpret the values produced by the systems. To this end, a decision support system capable of providing automated analysis of these data sets may help to improve the standards of care within maternity units and labour wards.

Ayres-de-Campos et al. [8] collected fetal CTG data from 2126 births and automatically processed this data into 21 attributes using SisPorto. This data was then classified by three expert obstetricians, and a consensual classification label assigned to each sample corresponding to fetal state (either normal, suspect, or pathological).

A particular challenge in applying classifiers to this data set is that not only does it involve considering multiple classes, but also that the data set is heavily imbalanced — with 1655 instances of normal CTG results, 295 instances of suspect CTG results, and 176 instances of pathological CTG results. As mentioned in Section II-B imbalanced data sets, and particularly those with multiple classes, introduce many difficulties in the training of machine learning algorithms [50]. This is compounded by the fact that many of the data augmentation techniques that are effective in binary classification problems can do more harm than good when applied to multi-class problems [3].

The multi-objective approach to evolving ANNs discussed in this paper overcomes the class imbalance problem by focussing on the simultaneous optimisation of classification accuracies for each class. This approach means that no assumptions about misclassification costs for each class are necessary, and information about the trade-offs between classification objectives is presented to the decision maker.

B. Encoding the problem

In order to use an EA for the optimisation of ANN classifier's weights, biases, and topology, the problem must first be encoded into a real-valued chromosome. Solutions to the problem will then adhere to this chromosome structure, which is iteratively subjected to search and variation operators throughout the optimisation process, before being decoded for evaluation. Figure 2 illustrates the chromosome structure used to store the problem variables for the problem. This encoding supports an ANN with 3 output neurons, a maximum of 2 hidden layers, and 12 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 2, and the last hidden layer must contain a minimum of neurons equal to the number of input neurons as seen in Equation 3. 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 or bias is restricted to the same boundary shown in Equation 4.

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

HL1.NEURONS Variables: 1	**HL2.NEURONS Variables: 1**	**IL.WEIGHTS Variables: 288**
HL1.WEIGHTS Variables: 576	**HL2.WEIGHTS Variables: 72**	**HL1.BIAS Variables: 24**
HL2.BIAS Variables: 24	**OL.BIAS Variables: 3**	

Fig. 2: Encoded chromosome for the three-objective ANN consisting of a maximum of 2 hidden layers (HL), input layer (IL), 3 neurons on the output layer (OL), and associated biases, totalling to 989 variables.

$$b(HL) = \{x \in \mathbb{Z} \mid i \leq x \leq 2i\} \tag{3}$$

$$w = \{x \in \mathbb{R} \mid -5 \leq x \leq 5\} \tag{4}$$

The algorithmic approach for generating the parameter boundaries which solutions must adhere to has been taken from [44]. This algorithm requires an input of the maximum number of hidden layers, minimum number of neurons per hidden layer (typically set to the number of input neurons), and the number of output neurons.

For the ANN encoding used in this paper, each candidate solution contains 989 variables, with the first 2 variables defining the number of hidden layers and the number of neurons on each respectively, the following 288 variables defining the weights for the input layer, the following 576 for the first hidden layer, and the following 72 for the second and final hidden layer. The remainder of the variables are used to store the biases associated with these layers.

Regardless of the ANN topology of each candidate solution (defined by the first two genes of the encoded chromosome) the solutions chromosome will still store all 989 genes, including weights and biases which are not used to configure a particular solutions ANN. These unused variables will remain unexpressed in the phenotype until the first two genes allow them to manifest. Genes can remain dormant for many generations before they are activated, which introduces the interesting feature of atavism[3] into this problem.

At each function evaluation, a chromosome is decoded from its encoded state described in Figure 2 and used to instantiate an ANN. This ANN is then used to classify the training data and the objective information is extracted and used to assess the chromosome's fitness based on the ANN result set. The algorithm used to decode a chromosome which conforms to the encoded structure has been taken from [44].

C. Optimisation results and discussion

The ANN encoding described in Section IV-B was optimised using the CMA-PAES-HAGA algorithm described in Section III and the performance evaluated using the objectives shown in Table I. These objectives represent the classification accuracies for each of the three classes considered in this problem and the overall classification accuracy. Note that these have been converted into minimisation objectives for the purpose of optimisation. As highlighted in Section IV-A, a particular challenge in designing a classifier for this data set is the extreme class imbalance (with more than three quarters of the data set classed as normal).

TABLE I: Performance objectives

Objective 1	Classification accuracy for normal fetal state
Objective 2	Classification accuracy for suspect fetal state
Objective 3	Classification accuracy for pathological fetal state
Objective 4	Overall classification accuracy

The performance of the optimised classifier is compared with a standard feed-forward multilayer perceptron, a support vector machine classifier with RBF kernel, and a Random Forests classifier. Results presented in [43] have shown that, for the two class version of this problem (i.e. distinguishing normal fetal states from pathological fetal states), the performance of SVM and Random Forests was comparable (with Random Forests obtaining slightly better results). Standard feed-forward multilayer perceptron results (with some manual tuning of the ANN topology) are also shown here as a point of comparison with the completely automated multi-objective optimisation of both weights and topology performed by the CMA-PAES-HAGA algorithm.

Figure 3 presents the confusion matrix for the standard feed-forward multilayer perceptron classifier, figure 4 presents the confusion matrix for the support vector machine classifier, and figure 5 presents the confusion matrix for the Random Forests classifier. Table II presents a summary of the performance of each of these classifiers in terms of both the overall accuracy and the M score (equivalent to the multi-class area under the ROC curve) calculated using the multi-class receiver operator characteristics curve approach from [26].

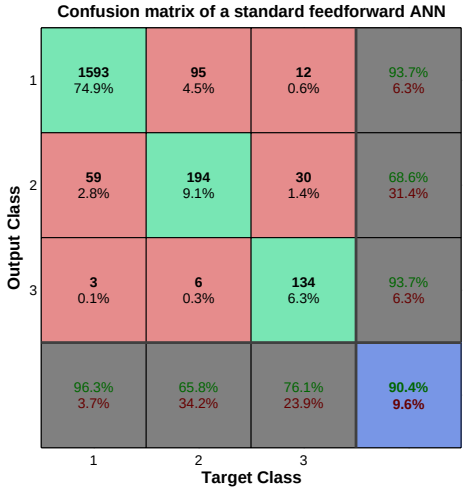

Fig. 3: Confusion matrix for the feed-forward multilayer perceptron ANN

Figure 6 shows a parallel coordinates plot of the trade-off surface of this classification problem. It can be seen that

[3] "Atavism is the tendency to revert to ancestral type. In biology, an atavism is an evolutionary throwback, such as traits reappearing which had disappeared generations before."

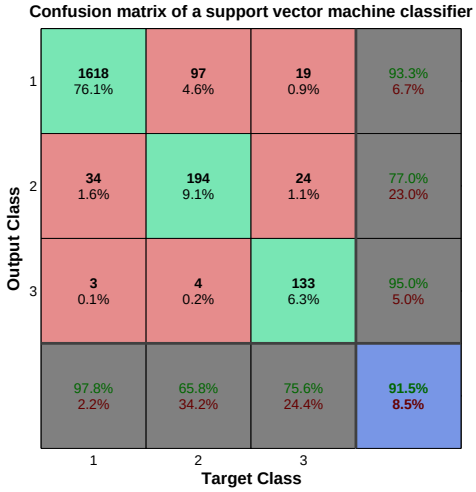

Fig. 4: Confusion matrix for the support vector machine classifier with RBF kernel

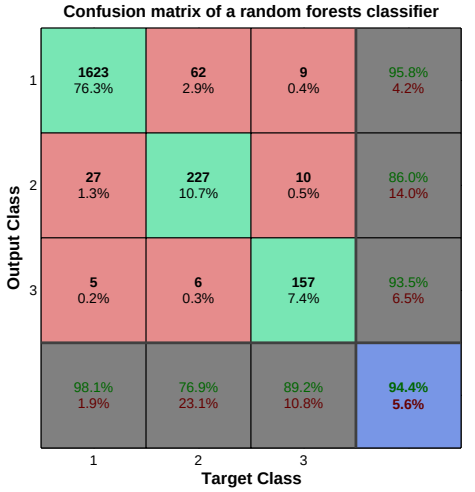

Fig. 5: Confusion matrix for the Random Forests classifier

there is a clear trade-off between objective 1 (classification accuracy of the normal fetal state) and objective 2 (classification accuracy of the suspect fetal state). Presenting the full trade-off surface (as in Figure 6) to a decision maker means that no *a priori* decisions have to be made about the costs of misclassification for each objective. In medical diagnosis, the costs of misclassification of each class are usually not equal — frequently the cost of misclassifying a sample from a minority class (representing an abnormality) as a majority class sample (representing the normal case) is highly expensive (for example, a patient may miss the opportunity for prompt and life saving treatment) [48].

TABLE II: Comparison of classification approaches

	Feed-forward multilayer perceptron	Support vector machine	Random Forests classifier
Accuracy	90.4%	91.5%	94.4%
M score	0.894	0.875	0.901

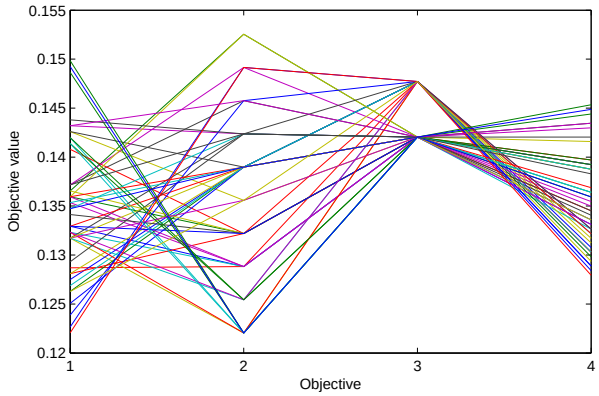

Fig. 6: Parallel coordinate plot of problem trade-off surface at the end of the optimisation process

The 50 candidate ANNs produced by the CMA-PAES-HAGA algorithm have also been evaluated using the multi-class M score from [26]. This resulted in a best M value of 0.913, a median M value of 0.911, and a standard deviation of 0.002. This indicates that the classifiers trained using the CMA-PAES-HAGA algorithm exhibit somewhat better discrimination compared to the benchmark classifiers shown in Table II and are more robust to the decision criteria used for deciding which class a sample belongs to. Figure 7 shows the confusion matrix for the best discriminating optimised classifier.

Comparing Figures 3, 4, and 5 to Figure 7 shows that whilst the other classifiers obtain a higher overall classification accuracy (90.4%, 91.5%, and 94.4% compared to 87.0% for the optimised ANN), the CMA-PAES-HAGA trained ANN obtains significantly better accuracy with respect to minority class recognition. It can be seen from Figures 3, 4, 5, and 7 that, although the misclassification rate of the normal class using the optimised multi-objective classifier is 12.9% compared to 3.7% for the standard feed-forward ANN, 2.2% for the support vector machine classifier, and 1.9% for the Random Forests classifier, the misclassification rates for the suspect class is significantly better using the CMA-PAES-HAGA trained ANN classifier (12.9% vs 34.2%, 34.2%, and 23.1%). Performance of the CMA-PAES-HAGA trained ANN classifier is significantly better than that of the standard multilayer perceptron and the support vector machine classifier for the pathological class (14.8% vs 23.9% and 24.4%), and roughly similar to that of the Random Forests classifier (14.8% vs 10.8%) with the CMA-PAES-HAGA trained ANN classifier less likely to mistake the pathological state for the normal fetal state but more likely to identify it as a suspect fetal state.

V. Conclusions

In this paper a novel method of optimising both the structure and parameters of an ANN using the state-of-the-art CMA-PAES-HAGA optimiser has been proposed. By using a multi-objective approach to considering trade-offs between the classification accuracies of each class in a multi-class classification problem, issues with class imbalance are addressed without the need for *a priori* integration of problem specific knowledge such as misclassification costs. The multi-

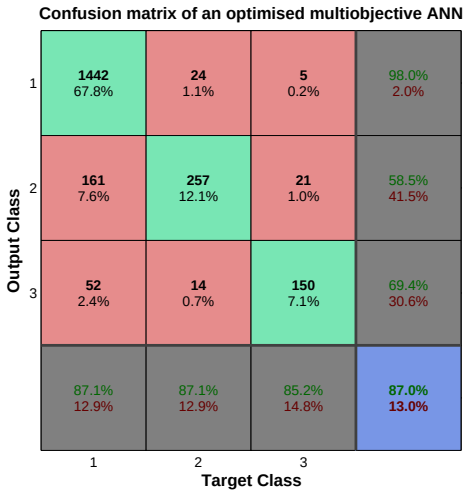

Fig. 7: Confusion matrix for the optimised multi-objective ANN

objective consideration of trade-offs between the accuracy of classification for each class in a problem means that there is no benefit in a classifier design inclining towards classifying the majority class correctly at the cost of mis-classifying one or more minority classes.

The multi-objective ANN optimisation approach proposed in this paper has been applied to the development of a decision support tool for the automated analysis of fetal cardiotocograms. A particular challenge in training classifiers with the data set used in this study is the extreme class imbalance, with more than 75% of the data belonging to the normal class. In this paper it is shown that, by handling the classification accuracies in a multi-objective way, it is possible to significantly improve minority class recognition for both the minority classes in this data set at the cost of slightly worse performance in terms of overall classification accuracy. This multi-objective approach has the added benefit that no *a priori* decisions have to be made about misclassification costs and the decision maker is presented with a number of trained ANNs with trade-offs that are well distributed across the Pareto front. The decision maker can then select an optimised solution which balances the number of false alarms with the number of cases where problems are missed.

REFERENCES

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

[2] H. A. Abbass, R. Sarker, and C. Newton. PDE: a pareto-frontier differential evolution approach for multi-objective optimization problems. In *Proceedings of the 2001 Congress on Evolutionary Computation*, volume 2, pages 971–978. IEEE, 2001.

[3] L. Abdi and S. Hashemi. To combat multi-class imbalanced problems by means of over-sampling techniques. *IEEE Transactions on Knowledge and Data Engineering*, 28(1):238–251, 2016.

[4] Z. Alfirevic, D. Devane, G. Gyte, et al. Continuous cardiotocography (ctg) as a form of electronic fetal monitoring (efm) for fetal assessment during labour. *Cochrane Database Syst Rev*, 3(3):CD006066, 2006.

[5] D. Arduini, G. Rizzo, and C. Romanini. Computerized analysis of fetal heart rate. *Journal of Perinatal Medicine-Official Journal of the WAPM*, 22(s1):22–27, 1994.

[6] M. Attik, L. Bougrain, and F. Alexandre. Neural network topology optimization. In *ICANN 2005: the 15th International Conference on Artificial Neural Networks*, pages 53–58. Springer, 2005.

[7] D. Ayres-de Campos, J. Bernardes, A. Costa-Pereira, and L. Pereira-Leite. Inconsistencies in classification by experts of cardiotocograms and subsequent clinical decision. *BJOG: An International Journal of Obstetrics & Gynaecology*, 106(12):1307–1310, 1999.

[8] D. Ayres-de Campos, J. Bernardes, A. Garrido, J. Marques-de Sa, and L. Pereira-Leite. Sisporto 2.0: a program for automated analysis of cardiotocograms. *Journal of Maternal-Fetal Medicine*, 9(5):311–318, 2000.

[9] M. J. Berry and G. Linoff. *Data mining techniques: for marketing, sales, and customer support*. John Wiley & Sons, Inc., 1997.

[10] A. Blum. *Neural networks in C++: an object-oriented framework for building connectionist systems*. John Wiley & Sons, Inc., 1992.

[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[12] L. Chen and D. Alahakoon. Neuroevolution of augmenting topologies with learning for data classification. In *Information and Automation, 2006. ICIA 2006. International Conference on*, pages 367–371. IEEE, 2006.

[13] G. S. Dawes, M. Moulden, and C. W. Redman. System 8000: computerized antenatal fhr analysis. *Journal of Perinatal Medicine-Official Journal of the WAPM*, 19(1-2):47–51, 1991.

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

[15] K. Deb, A. Pratap, S. Agarwel, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

[16] E. S. Draper, J. J. Kurinczuk, and S. Kenyon. MBRRACE-UK 2015 perinatal confidential enquiry : Term, singleton, normally-formed, antepartum stillbirth. Technical report, NHS England, 11 2015.

[17] C. Elkan. The foundations of cost-sensitive learning. In *International joint conference on artificial intelligence*, volume 17, pages 973–978. Lawrence Erlbaum Associates Ltd, 2001.

[18] 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.

[19] M. Farina and P. Amato. On the optimal solution definition for many-criteria optimization problems. In J. Keller and O. Nasraoui, editors, *Proceedings of the NAFIPS-FLINT International Conference 2002*, pages 233–238. IEEE Press, 2002.

[20] A. Fernández, V. López, M. Galar, M. J. Del Jesus, and F. Herrera. Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches. *Knowledge-based systems*, 42:97–110, 2013.

[21] 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. *IEEE Transactions on Neural Networks*, 21(5):750–770, 2010.

[22] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(4):463–484, 2012.

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

[24] 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.

[25] M. T. Hagan, H. B. Demuth, M. H. Beale, and O. D. Jesús. *Neural network design*. PWS Publishing, Boston, 1996.

[26] D. J. Hand and R. J. Till. A simple generalisation of the area under the roc curve for multiple class classification problems. *Machine learning*, 45(2):171–186, 2001.

[27] N. Hansen and A. Ostermeier. Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation. In *Proceedings of 1996 IEEE International Conference on Evolutionary Computation*, pages 312–317. IEEE, 1996.

[28] H. He, Y. Bai, E. A. Garcia, and S. Li. ADASYN: adaptive synthetic sampling approach for imbalanced learning. In *IEEE International Joint Conference on Neural Networks, 2008 (IEEE World Congress on Computational Intelligence)*, pages 1322–1328. IEEE, 2008.

[29] H. He and E. A. Garcia. Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, 21(9):1263–1284, 2009.

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

[31] 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.

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

[33] Y. Lin, Y. Lee, and G. Wahba. Support vector machines for classification in nonstandard situations. *Machine learning*, 46(1):191–202, 2002.

[34] K. Maeda. Computerized analysis of cardiotocograms and fetal movements. *Bailliere's clinical obstetrics and gynaecology*, 4(4):797–813, 1990.

[35] B. N. Manktelow, L. K. Smith, S. E. Seaton, P. Hyman-Taylor, J. J. Kurinczuk, D. J. Field, P. W. Smith, and E. S. Draper. MBRRACE-UK perinatal mortality surveillance report. Technical report, NHS England, 05 2016.

[36] O. Palomäki, T. Luukkaala, R. Luoto, and R. Tuimala. Intrapartum cardiotocography–the dilemma of interpretational variation. *Journal of perinatal medicine*, 34(4):298–302, 2006.

[37] 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.

[38] R. Reed. Pruning algorithms-a survey. *IEEE transactions on Neural Networks*, 4(5):740–747, 1993.

[39] R. Rifkin and A. Klautau. In defense of one-vs-all classification. *Journal of machine learning research*, 5:101–141, 2004.

[40] J. Rosser. Confidential enquiry into stillbirths and deaths in infancy (CESDI). highlights of the 5th annual report–part 3. *The practising midwife*, 1(12):30, 1998.

[41] S. Rostami and F. Neri. Covariance matrix adaptation pareto archived evolution strategy with hypervolume-sorted adaptive grid algorithm. *Integrated Computer-Aided Engineering*, 23(4):313–329, 2016.

[42] S. Rostami and F. Neri. A fast hypervolume driven selection mechanism for many-objective optimisation problems. *Swarm and Evolutionary Computation*, 2017. to appear.

[43] H. Sahin and A. Subasi. Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques. *Applied Soft Computing*, 33:231–238, 2015.

[44] A. Shenfield and S. Rostami. A multi objective approach to evolving artificial neural networks for coronary heart disease classification. In *2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pages 1–8, Aug 2015.

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

[46] Y. Sun, A. K. Wong, and M. S. Kamel. Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(04):687–719, 2009.

[47] 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.

[48] X. Wan, J. Liu, W. K. Cheung, and T. Tong. Learning to improve medical decision making from imbalanced data without a priori cost. *BMC Medical Informatics and Decision Making*, 14:111, 2014.

[49] 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.

[50] Z.-H. Zhou and X.-Y. Liu. Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Transactions on Knowledge and Data Engineering*, 18(1):63–77, 2006.

[51] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132, 2003.