Modified Decomposition Framework and Algorithm for Many-objective Topology and Weight Evolution of Neural Networks

Adham Salih
School of Mechanical Engineering
Tel-Aviv University
Tel-Aviv, Israel
adhamsalih@mail.tau.ac.il

Amiram Moshaiov
School of Mechanical Engineering
Tel-Aviv University
Tel-Aviv, Israel
moshaiov@tauex.tau.ac.il

2021 IEEE Congress on Evolutionary Computation (CEC) | 978-1-7281-8393-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/CEC45853.2021.9504953

***Abstract*—This paper presents a modified decomposition framework to support the Many-Objective Topology and Weight Evolution of Artificial Neural Networks (MaO-TWEANNs). Next, an algorithm, which is termed NEWS/D, is devised using the proposed framework. To validate its optimization capabilities, a numerical study is carried out. The performed numerical study includes demonstration problems ranging from three to seven objectives for which the ideal points are known. Finally, an additional numerical study is performed with respect to a possible real-life application. The latter study suggests that evolving class experts for multi-class classification problems could be enhanced using NEWS/D in a non-intuitive approach.**

Keywords— Artificial Neural Networks, Many-objective Optimization, Decomposition Approach, Neuro-evolution, Topology and Weight Evolution, Neural Architecture Search.

I. Introduction

Evolutionary algorithms are known to be useful for the design of Artificial Neural-Networks (ANNs). This approach is known as Neuro-Evolution (NE) [1]. Since the early 90s, the potential of NE has been successfully demonstrated in many studies involving application areas such as robotics, artificial life, computer games, and agent technologies [1, 2]. Unlike the traditional gradient-based search techniques, a major advantage of the evolutionary approach is its ability to simultaneously search for the optimal topology and weights of ANNs. This approach is termed TWEANNs (Topology and Weight Evolution of ANNs). Currently, most studies on TWEANNs are restricted to cases either with a single objective or with a few objectives (e.g., [3-7]).

Due to the numerical difficulties associated with an increasing number of objectives, studies have been dedicated to the development of Many-Objective Evolutionary Algorithms (MaOEAs). Such algorithms aim to find a good approximation of the Pareto-front and set for many-objective optimization problems [8]. Yet, there has been a lack of similar attempts to tailor MaOEAs to Many-Objective TWEANNs (MaO-TWEANNs). As suggested here, it is envisioned that the availability of such algorithms will open up new research and development possibilities.

When evolving topologies and weights, care must be taken while performing selection operators between individuals with topologies that were created at different generations. In other words, fair "age-based competitions" among topologies are a major concern when evolving both topologies and weights (e.g., [7]). As a result, modification to the decomposition framework of [9] is suggested to adapt it to MaO-TWEANNs. The attractiveness of using a decomposition approach is that it is a leading approach for dealing with many-objective optimization problems [10]. In addition, this approach allows biasing the search based on the preferences of objectives.

The proposed framework is used to devise a specific algorithm, which is hereby termed as Neuro Evolution of Weights and Structures by Decomposition (NEWS/D). The suggested algorithm is validated and compared. Furthermore, an initial study is conducted on an envisioned application.

The rest of this paper is organized as follows. Section II outlines the background material for this study, whereas section III introduces the proposed framework for MaO-TWEANN. Next, in section IV, the devised algorithm is described. Section V presents the validation and comparison produces, while section VI provides the numerical results and their analyses. Finally, section VII summarizes and concludes this study.

II. Background

A. Multi- vs. Many-objective Optimization

Consider the following vector optimization problem:

$$\min_{x \in X} \quad [F_1(\boldsymbol{x}), \dots, F_M(\boldsymbol{x})] \tag{1}$$

where M is the number of objective-functions to be minimized, and $X \subseteq \mathbb{R}^d$ is the set of feasible solutions. According to the Pareto-optimality approach, solving this problem provides the Pareto-optimal-set of non-dominated solutions and their associated performance-vectors, which constitute the Pareto-front. The following provides some definitions of the Pareto optimality concept. A solution $\boldsymbol{a}$ with a performance vector $\boldsymbol{F}(\boldsymbol{a}) = [F_1(\boldsymbol{a}), \dots, F_M(\boldsymbol{a})]$ is said to dominate a solution $\boldsymbol{b}$ with a performance vector $\boldsymbol{F}(\boldsymbol{b}) = [F_1(\boldsymbol{b}), \dots, F_M(\boldsymbol{b})]$, denoted by $\boldsymbol{a} \preccurlyeq \boldsymbol{b}$, if the following condition is met:

$$\forall i \in \{1, \dots, M\}, F_i(\boldsymbol{a}) \le F_i(\boldsymbol{b}) \;\&\; \exists i \text{ s.t. } F_i(\boldsymbol{a}) < F_i(\boldsymbol{b}) \tag{2}$$

For a given optimization problem, the Pareto-optimal-set is the set of all non-dominated feasible solutions of the problem. Namely, a solution $\boldsymbol{a} \in X$ is called Pareto-optimal with respect to X if and only if there is no solution $\boldsymbol{b} \in X$ for which $\boldsymbol{b} \preccurlyeq \boldsymbol{a}$. Solving (1) means finding the Pareto-optimal-set and the associated Pareto-front of the problem.

Commonly, there is no analytical solution to the above problem. Hence, numerical techniques are needed to get at least a good approximation for the Pareto-optimal-set which has performances as close as possible to the true Pareto-front. Over the last two decades, a set of reliable Multi-Objective Evolutionary Algorithms (MOEAs) has been developed for solving the problem in (1).

As pointed out in [8], traditional MOEAs may fail to converge under an increasing number of objectives. To distinguish between such algorithms and those that aim at solving problems with an increasing number of objectives, the term Many-objective Optimization Problems (MaOPs) and MaOEAs have been introduced. Using "many" as opposed to "multi" commonly refer to solving problems with more than three objectives. As apparent from reviews such as in [8], the development of MaOEAs has reached a certain maturity. A popular evolutionary approach to deal with MaOPs is the decomposition-based approach [9, 10], which is modified here for NE. As declared in the introduction section, there are two main advantages when using this approach including a. the search can be biased according to the objective preferences; b. it can handle an increasing number of objectives.

B. *Topology and Weight Evolution of ANN*

Algorithms for single-objective TWEANNs, such as NEAT [7], have been studied over more than two decades. In contrast, attempts to develop such algorithms for multi-objective TWEANNs have been relatively scarce. The most prevailing versions combine NEAT with NSGA-II [11]. However, NSGA-II is known to not being suitable for solving MaOPs. Hence, to cope with MaO-TWEANNs, it appears logical to replace NSGA-II with an alternative MaOEA to be combined with NEAT or with any other algorithm for single-objective TWEANNs.

It appears that most studies on algorithms for multi-objective TWEANNs are based on MOEAs which are known to be restricted to dealing with a few objectives. To the best of our knowledge, the only exceptions are the studies in [6, 12, 13]. However, these algorithms were tested using problems with four objectives. Here, the testing is done with a set of MaOPs of different dimensionality ranging from three to seven objectives. In addition, the algorithm, which is proposed here, has the advantages that are inherent to the use of the decomposition approach (see section II.A).

III. The Proposed Framework

A. *Framework Description*

A framework for MaO-TWEANNs is introduced, which is based on the decomposition technique for MaOEAs [9]. In contrast to the decomposition-based evolutionary framework of [9], here the proposed evolutionary framework is tailored to TWEANNs. According to [7], when a new topology is generated, it is expected to have low performance as compared with that of a mature topology. This is due to the lack of sufficient tunning of the weights of the new topology. Hence, special care is taken to ensure that no competition between novice and mature topologies occurs. The evolutionary framework for the development of decomposition-based algorithms for MaO-TWEANNs is outlined in Framework 1.

Framework 1: MaO-TWEANNs	
Input	*Algorithm's parameters*
Output	*Archive*
Initialize	*Population, Archive, and Topology Record*
1	Perform **Sub-problem Generation Procedure**
2	**While** stopping criteria is not met:
3	Perform **Score Assignment Procedure**
4	Perform **Elite Set Procedure**
5	Perform **Protected Set Procedure**
6	Perform **Selection Procedure**
7	Perform **Reproduction Procedure**
8	Perform **Population Update Procedure**
9	Update the Archive
10	$gen \leftarrow gen + 1$
11	**End While**

The following briefly describes the associated procedures and their goals. It should be noted that the description refers to a framework for the creation of alternative algorithms, rather than to describe specific procedures. For a specific algorithm, which is based on the proposed framework, the reader is referred to section IV.

The Sub-problem Generation Procedure aims to realize the main idea of the decomposition approach. It involves the generation of a set of K single-objective sub-problems based on the requirements of the users. The first procedure of the generation loop is the Score Assignment Procedure, which assigns a score vector $\boldsymbol{s_i} \in \mathbb{R}^K$ for each i^{th} individual of the population. The k^{th} score, $k \in \{1,2,\ldots,K\}$, is the calculated fitness of the individual relating to the scalarized objective of the k^{th} sub-problem. This is followed by the Elite Set Procedure. In this procedure, for each of the k^{th} sub-problem, a set of elite individuals is created based on their k^{th} scores. These sets, from all sub-problems, are to be united into an elite set, which aims to avoid loss of performances over the generations.

Next, the Protected Set Procedure is performed. In this procedure, a protected set of individuals is created, which includes individuals of the population that have novice topologies. This is followed by the Selection Procedure per sub-problem using the sub-problem scalar score. The Reproduction Procedure is to be done using a multi-set of all the selected individuals from all the sub-problems as well as those of the elite and the protected set. Before the next generation, the Population Update Procedure is applied. The population is updated by replacing some of its individuals with protected and elite individuals. The archive is updated by a. adding offspring that are non-dominated, and b. removing dominated solutions. It is noted that when the stopping condition is met, the archive includes the best solutions of the run.

B. Working principles

This section aims to discuss the working principles of the proposed framework including the difficulties of merging TWEANN algorithms with MaOEA and especially decomposition-based types. The first principle is the “protection principle”. This principle is commonly applied in TWEANN algorithms to ensure age-based competitions [7]. In the proposed framework, such competitions are achieved as follows. First, a set of protected individuals is created using the Protected Set Procedure. This set includes solutions from all novice topologies. Then, if a protected topology fails selection, it is restored from the protected set into the evolved population. It is noted that different conditions could be used for the protection rule. For example, a topology might be considered as a novice when a pre-defined number of generations has not been reached since its creation. The second principle of the suggested framework is the sub-problem elitist selection. During the selection procedure, the individual which has the best score in a particular sub-problem is automatically selected for the reproduction procedure. The third principle is the diversity principle. This means that the selection mechanism must ensure a diverse set of solutions for reproduction. In the proposed framework, in each selection step, a sub-problem is randomly chosen. Then, the individual selection is performed based on its score regarding the chosen sub-problem. This encourages diversity by way of selecting an individual per each of the chosen problems. Diversity is further strengthened by the elitist selection, which ensures that each sub-problem has its best individual in the general population.

IV. ALGORITHM DESIGN

Based on the proposed framework, a specific algorithm, which is hereby termed as Neuro Evolution of Weights and Structures by Decomposition (NEWS/D), has been devised. The following presents the specific details of NEWS/D. The main procedure, which is described in Algorithm 1, follows the main steps of the proposed framework.

Algorithm 1: NEWS/D main procedure	
Input	N, K, MAX_{NT}, gen_{TM} , *Parameters of the Stopping Criterion*
Output	Archive
Initialize	*Population* ($P_{gen=0}$), *Archive* and *Topology Record* (TR)
Steps 1-11	Follows Framework 1

The following inputs are provided by the user. The first input is N, which is the number of individuals in the population. The second input is K, which is the number of sub-problems. The third input is a pre-defined number MAX_{NT}, which states the maximal number of evolving topologies in each generation. It aims to balance between the topology search and the weight search to control the search resources. As explained in the Selection Procedure below, the condition $N > K + MAX_{NT}$ should be satisfied when the above inputs are selected. The fourth input is the number of protected generations, gen_{TM}, which is a pre-defined number of generations. It is used to determine if the topology is to be considered as a novice. In the proposed NEWS/D algorithm, a topology is considered as matured when the number of generations, since its creation, has reached the pre-defined number of generations, gen_{TM}. The user should also define the parameters that are needed for the stopping criterion. Assuming it involves the maximal number of generations (gen_{max}), then $gen_{TM} \ll gen_{max}$.

During the initialization stage, an initial population $P_{gen=0}$ of networks is created. In particular, the encoding scheme of NEAT, without speciation, is used in the current implementation. This appears sufficient for the presented demonstrations of the proposed framework. In such an encoding scheme, each ANN is represented using two lists. The first list is the nodes' gene list, which describes the nodes’ input signals, the nodes' output signals, and the nodes’ type (input, output, or hidden neuron). The second list describes the network connections. For each connection, the list includes indices of the connected nodes, the weight of the connection, enabling bit, and an innovation number. Further details on the encoding scheme and the related reproduction mechanisms can be found in [7]. In general, the initial population is set with the smallest possible topology and with random weights.

The generation loop of the main procedure aims to update an archive of the non-dominated set of solutions. At the end of the evolutionary process, the archive contains the approximated set of solutions. For the implementation of NEWS/D, for each of the evolved topologies, a record is set. It includes a. the structure code, b. the generation of creation, c. the current number of individuals with the topology, and d. the protection status. For each change in the topologies, the elements of this record are updated.

The main procedure of NEWS/D applies secondary procedures, as listed in Framework 1. First, Sub-problem Generation is applied using the Tchebycheff (TCH) approach as in [9]. It should be noted that this approach, as opposed to the weighted-sum approach, allows solving problems in which the front is either convex or non-convex. The considered sub-problems are defined based on a set $\lambda = \{\boldsymbol{\lambda}^1, \boldsymbol{\lambda}^2, \dots, \boldsymbol{\lambda}^K\}$ of K evenly spread weight vectors, where each of these vectors is associated with one of the sub-problems. Let X be the set of all feasible solutions and let $\boldsymbol{a}, \boldsymbol{b} \in X$. Also, $F_i(\boldsymbol{a})$ and $F_i(\boldsymbol{b})$ are the performances in the i^{th} objective of $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively. Using the TCH approach, the k^{th} sub-problem is defined as:

$$\min_{\boldsymbol{a}\in X} g^{te}(\boldsymbol{a}|\boldsymbol{\lambda}^k, \boldsymbol{z}^*) = \min_{\boldsymbol{a}\in X} \max_{1\le i\le M} \left\{\lambda_i^k |F_i(\boldsymbol{a}) - z_i^*|\right\} \\ z_i^* = \min_{\boldsymbol{b}\in X} F_i(\boldsymbol{b}) \tag{3}$$

where $g^{te}(\boldsymbol{a}|\boldsymbol{\lambda}^k, \boldsymbol{z}^*)$ is the score of solution $\boldsymbol{a}$ in the k^{th} sub-problem score associated with the weight vector $\boldsymbol{\lambda}^k$, and z_i^* is the i^{th} component of the ideal point. It should be noted that the above definition is for a minimization problem. In the case of a maximization problem, the solution should be based on the use of $z_i^* = \max_{\boldsymbol{b}\in X} F_i(\boldsymbol{b})$. The procedure for generating the sub-problem in the current implementation, which has been adopted from [14], a different set of sub-problems might be created at each run.

During the generation loop of NEWS/D, the Score Assignment Procedure aims to assign a score to each individual (network) of the population P_{gen}. The score is $g^{te}(\boldsymbol{a}|\boldsymbol{\lambda}^k, \boldsymbol{z}^*)$, as defined in (3), where $\boldsymbol{a} \in P_{gen}$ and $\boldsymbol{b} \in (P_{gen} \cup \text{Archive})$. Next, the Elite Set Procedure is applied. It finds the best individual for each of the sub-problems. Furthermore, these individuals are used to create the elite set. It is noted that the created set is not a multi-set. Namely, repeated individuals are not included.

Next, the Protected Set Procedure is carried out. It sets the topology protection status for each of the topologies in the current population. For each protected topology, it saves a randomly selected individual to form a set of protected individuals (PI).

Algorithm 2 gives the pseudo-code of the Selection Procedure. As explained in the introduction, the neighborhood-based selection procedure of the original decomposition framework of [9] is not suitable for TWEANNs. For this reason, the current selection procedure avoids the use of subpopulations. Here, the elite set and the protected set are united. Their union forms a set of selected individuals. This set is to be expanded by an iterative selection procedure until $|SI| = N$, assuming that $N > K + MAX_{NT}$.

Algorithm 2: Selection Procedure	
Input	$PI, E(Elite), P_{gen}$
Output	SI (Selected Individuals)
Initialize	$SI = \emptyset$
	$SI \leftarrow PI \cup E$
	While $|SI| < N$
	Randomly select a sub-problem k
	Randomly select two individuals $p_1, p_2 \in P_{gen}$
	If p_1 is better than p_2 in sub-problem k
	Add p_1 to SI
	Otherwise
	Add p_2 to SI
	End If
	End While

The Reproduction Procedure includes two types of reproduction mechanisms. The first type is for topology reproducing and the second is for weight updating. Here, the NEAT reproduction scheme is used for the topology reproduction following [7]. For the second type, the weight vectors are updated as described in Section VI.

At the end of each generation, the population is updated. The update includes two main steps. In the first step, elitism is applied for each sub-problem. The second step ensures that the updated population includes individuals with protected topologies. Finally, after each generation, the Archive is updated.

V. VALIDATION AND COMPARISON PROCEDURES

The optimization capabilities of the proposed NEWS/D are tested using data from multi-class classification datasets. Yet, it is important to note that while based on classification data, the evaluation aims to check the optimization characteristics of the algorithms, rather than generalization or learning. The main advantage of using multi-class classification datasets is that the ideal point of each such problem is known, which helps to evaluate the optimization capabilities of the tested MaO-TWEANN algorithm.

A. Problem Definition

Let a multi-class classification dataset be with N_C classes and N_s samples. Let, $c \in \{1, \dots, N_C\}$ be the class index and let $i \in \{1, \dots, N_s\}$ be the sample index. Let $\boldsymbol{x_i}$ be the input (feature) vector of the i^{th} sample and let $y_i^c \in \{0,1\}$ be the value of the c^{th} component of the desired output vector of the i^{th} sample. For samples belonging to the c^{th} class, y_i^c is equal to 1. Also, let $\tilde{y}_i^c \in \mathbb{R}$ be the value of the c^{th} component of the output vector, as obtained by feeding the input $\boldsymbol{x}_i$ to the neural network. The proposed evaluation method is based on the minimization of the objective vector $\boldsymbol{F} \in \mathbb{R}^{N_c}$, where the c^{th} component of $\boldsymbol{F}$ is given by:

$$F_c = \frac{1}{2N_s} \sum_{i=1}^{N_s} (\tilde{y}_i^c - y_i^c)^2 \tag{4}$$

The numerical study includes eight multi-class datasets. Table I, which is based on [15], summarizes the used datasets. It is noted that by the problem definition, the ideal point of the Pareto-front is known for each of the benchmark problems. Since optimization rather than learning is done here, for each of the eight benchmarks, the complete datasets are used. These sets serve as references for calculating the error in (4).

TABLE I: DATASETS CHARACTERISTICS

Dataset	#Classes	#Instances	#Features
Wine	3	178	13
Seeds	3	210	7
Vehicle	4	846	18
UKM	4	403	5
BT	6	106	9
Glass	7	214	9
Zoo	7	101	16
Segment	7	2310	19

B. Evaluation Procedures

As noted above, only the ideal points are known, rather than the true Pareto-fronts of the optimization problem. Hence, to evaluate the algorithm, two types of evaluation procedures are carried out, absolute and relative evaluation procedures are suggested as detailed in the following subsections. First, the tested algorithm is evaluated with respect to the true ideal point (absolute evaluation). Next, to test the ability of the algorithm to find a good approximation of the Pareto-set, relative evaluation procedures are suggested.

a) Absolute Evaluation Procedure

The procedure compares the ideal point, as found by the tested algorithm, with the true one. The absolute evaluation procedure is based on the following ideal point absolute error measure, where N_C is the number of classes in the considered problem. The error formulation results that the true ideal point of the considered optimization problem is at the origin of the objective space. It is noted that, in (5), z_i is the normalized value of the i^{th} component of the ideal point, as obtained by the evaluated algorithm.

$$err = \frac{1}{N_C}\sum_{i=1}^{N_C} z_i \tag{5}$$

b) Relative Evaluation Procedures

For the relative evaluations, it is proposed to compare the results of running the tested algorithm on the aforementioned optimization problem with those obtained by two different single-objective algorithms. The suggested comparison techniques are hereby referred to as Technique A and B. Technique A involves the use of a selected algorithm for single-objective TWEANNs, which is hereby referred to as Algorithm A. In contrast, Technique B involves using a selected algorithm for single-objective NE with fixed topology, which is hereby referred to as Algorithm B. The suggested comparison techniques are detailed in the following.

The main idea behind the evaluation techniques is to apply Algorithm A and B on sub-problems that were evaluated by the tested algorithm. With this respect, it is noted that NEWS/D generates different sets of sub-problems at different runs. Hence, in this case, a record of the generated sub-problems is maintained for each run. This record is used to define the single-objective problems to be solved by Algorithm A and B for making comparisons.

Given the stochastic nature of evolutionary algorithms, multiple runs must be used to make any statistical inference. Recall that the tested algorithm uses K sub-problems per each run. Applying the tested algorithm N_r times on each of the benchmark problems, the total number of possible comparisons per problem, for each technique, is KxN_r. For practical reasons, the user may decide to use a subset of these options for the comparisons. In the current study, the entire set of possible comparisons is used.

Applying Algorithm A on the selected KxN_r sub-problems will produce a neural network, NN_{kr}^A, for the k^{th} sub-problem of the r^{th} run. Similarly, using the algorithm of Technique B, a neural network, NN_{kr}^B, is obtained for the k^{th} sub-problem of the r^{th} run. Recall that Technique B is based on using a selected algorithm for single-objective NE with fixed topology. In this case, the employed topology for the k^{th} sub-problem of the r^{th} run is selected as the best topology obtained by the tested algorithm for that sub-problem by that run.

Given the multiplicity of possible comparisons, a worst-case approach is suggested as follows. This approach is based on finding the sub-problem, per each run, in which the tested algorithm has the worst performance in comparison with that of the applied algorithm A or B. For each technique A and B, and for each run, a representative sub-problem is found, and the associated representative performances are calculated as follows. First, let $s_k^{MaO}(r)$ be the k^{th} sub-problem score as obtained by the tested algorithm in the r^{th} run. Also, let $s_k^A(r)$ and $s_k^B(r)$ be the corresponding scores, as obtained by NN_{kr}^A and NN_{kr}^B, respectively. Next, for the r^{th} run and for each of the T^{th} technique, where $T \in \{A, B\}$, let the representative sub-problem $k^T(r)$, be calculated according to the worst-case principle as follows:

$$k^T(r) = \underset{k\in\{1,\ldots,K\}}{\operatorname{argmin}}\{s_k^{MaO}(r) - s_k^T(r)\} \tag{6}$$

It should be noted that the associated representative performance of the T^{th} technique is $s_{k^T(r)}^T(r)$, while that of the tested algorithm is $s_{k^T(r)}^{MaO}(r)$. Finally, these representative performances, as obtained over the runs, are used for making statistical inferences on the relative capabilities of the algorithms.

In addition, it is suggested to perform comparisons that are based on front measures. In this study, the Inverted Generational Distance (IGD) measure is used to make a statistical inference, as commonly done in studies on multi-objective optimization (e.g. [9]). For the calculations of measures such as the IGD, it is proposed to use the non-dominated set of solutions of the union of all solutions and its performance vectors, as obtained by uniting and sorting the solutions from all runs of the tested algorithm and of the single-objective algorithms.

Recall that for each run of the tested algorithm there is an obtained approximated Pareto optimal set and front. In order to compare it with the results of any of the single-objective algorithms, these algorithms are run multiple times according to the sub-problems of the run of the tested algorithm. The obtained solutions by each of the single-objective algorithms are used to create an approximated front per each of these algorithms. Finally, per each run of the tested algorithm, the IGD value and the ideal point of the obtained front are compared with the values of the associated fronts of the single-objective algorithms.

VI. Experimental Setup and Results

The following experimental study was conducted to evaluate NEWS/D using the datasets of Table I. Given the stochastic nature of evolutionary algorithms, for each of the datasets, 31 runs of the proposed NEWS/D were conducted. In this study, Algorithm A and Algorithm B were used. These are described in section (V.B.b).

Table II summarizes the parameters, which were used in these runs, where N_C is the number of classes in a particular problem. In the current study, the stopping criterion of NEWS/D is a total number of generations, which is denoted as gen_{max}. In order to perform fair comparisons, the number of evaluations in each run of the single-objective algorithms was limited. This was done with a population size of $\lceil N/\sqrt{K}\,\rceil$ for the single-objective runs and a with $\lceil gen_{max}/\sqrt{K}\rceil$ as the generation limit.

For the comparison by Technique A, algorithm A was chosen as a single-objective version of NEWS/D. Namely, the sub-problem generator has been replaced with a single sub-problem. For the comparison by Technique B, algorithm B was a version of the single-objective algorithm, with a de-activation of the topology search.

TABLE II: RUN PARAMETERS OF NEWS/D

Parameter	Value
Num. of Generations gen_{max}	$40 \cdot N_C$
Population size N	$40 \cdot (N_C - 1)$
Number of sub-problems K	$\lceil N/1.5 \rceil$
Number of protected generations gen_{TM}	$\lceil gen_{max}/ 20 \rceil$
Maximum number of topologies MAX_{NT}	$\lceil N/10 \rceil$
Topology crossover probability	0.8
Add node probability	0.2
Add connection probability	0.1
Weight Mutation (WM) mechanism	Polynomial
WM parameters (probability, distribution index)	(0.2,15)
Weight Crossover (WC) mechanism	SBX
WC parameters (probability, distribution index)	(0.8,20)

It should be noted that the selection of these single-objective algorithms for the relative comparisons is done because it helps to ensure that the search parameters are the same. It is also noted that the absolute accuracy of each of the selected single-objective algorithms is checked as detailed below. In all experiments, when evolving topologies, the initial population included ANNs with a minimal topology and random weights. In these networks, the number of input neurons was defined according to the number of features in the datasets, and the number of outputs corresponded to the number of classes. For the activation function, the sigmoid function was selected for all neurons. In order to make statistical inferences, the Wilcoxon test with a significance level of 5% is used for all comparisons.

A. Algorithms' Validation by the Ideal Points

Table III presents a summary of the medians and the standard deviation of the ideal point errors (see equation (5)). The results of NEWS/D in Table III indicate its (absolute) ability to find a good approximation of the true ideal point. In addition, the absolute ideal point evaluation is carried out for each one of the single-objective algorithms. The absolute results in Table III suggest that the three algorithms produced a good approximation of the ideal point.

TABLE III: ABSOLUTE EVALUATION RESULTS – IDEAL POINT

Dataset	NEWS/D	Algorithm-A	Algorithm-B
Wine	**1.25e-3(±6.25e-4)**	8.08e-3(±1.01e-3)	7.69e-3(±1.05e-3)
Seeds	**5.95e-3(±1.00e-3)**	1.13e-2(±1.25e-3)	1.13e-2(±1.25e-3)
Vehicle	**3.06e-2(±1.46e-3)**	3.96e-2(±9.42e-4)	3.93e-2(±1.41e-3)
UKM	**2.26e-2(±1.87e-3)**	2.88e-2(±1.16e-3)	2.79e-2(±1.02e-3)
BT	**1.47e-2(±3.26e-4)**	1.89e-2(±7.27e-4)	1.93e-2(±8.13e-4)
Glass	**1.08e-2(±4.10e-4)**	2.09e-2(±5.09e-4)	2.09e-2(±4.97e-4)
Zoo	**1.05e-3(±4.62e-4)**	4.89e-3(±7.37e-4)	5.27e-3(±7.15e-4)
Segment	**8.36e-3(±6.79e-4)**	1.57e-2(±1.61e-3)	1.63e-2(±1.23e-3)

The absolute results of algorithms A and B confirm that they may be used for the relative evaluations of NEWS/D. It should be noted that based on the Wilcoxon test (not reported here), it can be concluded that the proposed algorithm produces significantly better results for all the ideal points of the benchmarks in comparison with each of the single-objective algorithms.

B. Relative Comparison Results

Table IV presents a summary of the medians and the standard deviations of the representative performances (see section V.B.b) as obtained by each of the relative comparison techniques A and B, respectively. It should be noted that the proposed NEWS/D provided better median and standard deviation for each of the benchmarks, as compared with both algorithms A and B.

The results of using the Wilcoxon test for each comparison technique indicate that, in general, and for the worst-case scenario, the proposed algorithm produces significantly better results in most of the benchmarks. The obtained p-values are listed in the tables, the underlined values represent tests with no significant results, based on the 5% criterion. When making this conclusion, it should be recalled and emphasized that the representative performances are evaluated using a worst-case approach in favor of the single-objective algorithms.

TABLE IV: COMPARISON RESULTS – TECHNIQUE A AND B

Dataset	Median of $s_{k^A}^{MaO}$	Median of $s_{k^A}^{A}$	p-value
Wine	**2.24e-07(±6.54e-05)**	1.71e-03(±1.06e-03)	7.1e-11
Seeds	**1.96e-07(±1.80e-04)**	1.64e-03(±8.78e-04)	7.8e-11
Vehicle	5.73e-03(±3.17e-03)	**5.09e-03(±3.66e-03)**	<u>6.1e-01</u>
UKM	2.23e-03(±2.28e-03)	**2.21e-03(±2.92e-08)**	<u>1.1e-01</u>
BT	**1.41e-03(±4.94e-04)**	3.63e-03(±6.21e-04)	6.5e-11
Glass	**5.71e-04(±1.84e-04)**	1.58e-03(±3.96e-04)	2.3e-11
Zoo	**5.08e-04(±5.37e-04)**	3.10e-03(±4.91e-04)	1.4e-11
Segment	**1.26e-04(±5.23e-04)**	5.47e-03(±7.19e-04)	1.4e-11
Dataset	Median of $s_{k^B}^{MaO}$	Median of $s_{k^B}^{B}$	p-value
Wine	**2.18e-07(±3.46e-07)**	1.32e-03(±7.68e-04)	1.4e-11
Seeds	**2.12e-07(±1.54e-07)**	1.62e-03(±8.78e-04)	1.4e-11
Vehicle	**6.04e-03(±2.49e-03)**	7.10e-03(±2.84e-03)	1.9e-02
UKM	**2.87e-07(±2.44e-03)**	1.80e-03(±2.23e-03)	1.5e-04
BT	**1.53e-03(±5.68e-04)**	3.22e-03(±5.76e-04)	6.5e-11
Glass	**5.28e-04(±1.21e-04)**	9.21e-04(±2.11e-04)	4.3e-09
Zoo	**6.19e-04(±5.56e-04)**	2.48e-03(±3.52e-04)	1.9e-11
Segment	**1.16e-04(±5.36e-04)**	4.11e-03(±8.02e-04)	1.7e-11

Table V presents the medians and the standard-deviation of the IGD measure as obtained by NEWS/D and by each of the single-objective algorithms. As evident from the table, NEWS/D obtained better median and standard deviations for each benchmark. The results of using the Wilcoxon test to compare the proposed algorithm with algorithm A and B indicate that the proposed algorithm produces significantly better IGD results for each of the benchmarks. The maximal obtained p-value is 1.4e-11.

TABLE V: COMPARISON RESULTS – IGD

Dataset	NEWS/D	Algorithm-A	Algorithm-B
Wine	**9.07e-3(±2.61e-3)**	4.63e-2(±1.14e-3)	6.14e-2(±1.23e-3)
Seeds	**8.33e-3(±1.70e-3)**	3.08e-2(±5.58e-3)	4.04e-2(±7.64e-3)
Vehicle	**1.39e-2(±1.57e-3)**	2.48e-2(±1.26e-3)	2.63e-2(±1.52e-3)
UKM	**1.58e-2(±1.89e-3)**	3.11e-2(±1.49e-3)	3.38e-2(±1.38e-3)
BT	**2.10e-2(±1.64e-3)**	4.33e-2(±1.40e-3)	4.27e-2(±1.68e-3)
Glass	**2.17e-2(±2.68e-3)**	4.07e-2(±1.42e-3)	3.79e-2(±1.04e-3)
Zoo	**9.41e-3(±1.48e-3)**	4.19e-2(±3.33e-3)	4.22e-2(±3.34e-3)
Segment	**1.93e-2(±1.27e-3)**	5.02e-2(±1.67e-3)	4.80e-2(±1.53e-3)

C. *Objective Conflict*

The plots in Fig 1 present, for each dataset, the error trade-offs in the obtained approximation of the Pareto-front.

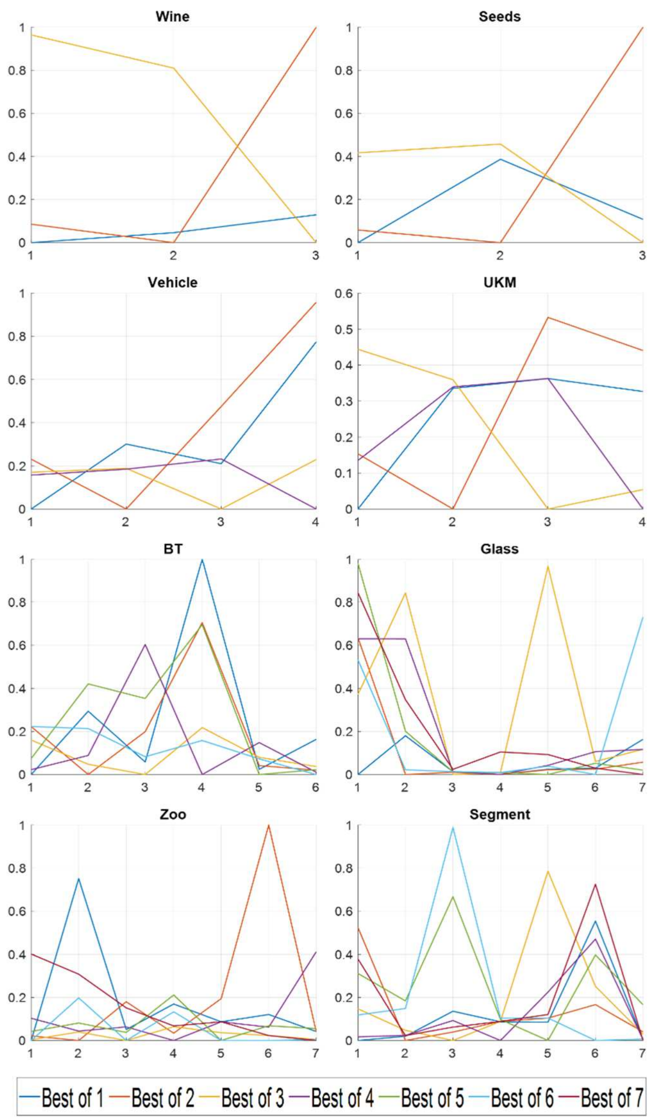

Fig. 1. Error Trade-offs between Classes

Each plot provides a parallel coordinate illustration of performances of the best solutions per each class. Each line connects the performances of an individual over the objectives. The y-axis is the normalized performance values, and the x-axis is the objective (class) indices. As evident from these plots, most of the class errors conflict, which results in a front. It is noted that, for all datasets, the obtained set of solutions includes more than one topology. For example, for the BT dataset, the obtained set includes 111 different topologies, whereas, in the Glass dataset, 213 topologies have been observed.

D. *Towards a Real-life Application*

This section provides a numerical study towards a real-life application. Consider an envisioned ensemble of expert classifiers that aims to solve a multi-class classification problem. Each expert member of the ensemble will aim to detect if a given input belongs to a particular class or not. Each of the envisioned experts will be designed based on a Deep Neural Network (DNN). Each of the networks is expected to have a mutual DNN part that will be designed to extract high-level features from the raw inputs. This mutual part of the DNNs will be accompanied by a Unique Expert Network (UEN) per each expert. Each of the UENs will be trained to map the high-level features, as obtained from the mutual part, into an output that indicates to what a degree the raw input belongs to the considered expert class. It is envisioned that using NEWS/D will support the design of the envisioned UENs.

In the context of UENs, the aim here is to find the best network for each component of the objective vector for the problem presented in Section V.A, rather than to find the entire set of approximated Pareto-optimal networks. The following provides a numerical study that aims to compare two NEWS/D-based search methods. In the 1st method, which is non-intuitive, the sought optimal networks are found by solving sub-problems that are based on weight vectors that are uniformly distributed over the entire objective space. In contrast, the 2nd method, which is the intuitive one, is based on running NEWS/D only with the sub-problems that are associated with binary weight vectors in which only one component is one and the rest are zeros. These methods are denoted here as F-NEWS/D and NF-NEWS/D, respectively. The run parameters are as in Table II except for the NF-NEWS/D in which the number of sub-problems was set to the number of the classes in the considered multi-class dataset.

While using optimization, rather than generalization, it is still expected that the conclusions from the following study are indicative also to the future modification of NEWS/D for the development and training of the UENs.

The numerical study includes 41 comparisons, one per class component per each of the eight problems that are based on Table 1. Table VI shows some typical results including the median and the standard deviation as obtained by each method. For each comparison, the resulting p-value, by applying the rank-sum Wilcoxon test, is also listed in the table. The results from the entire 41 comparisons, which are partially shown here, indicate that F-NEWS/D has been superior to the intuitive NF-

NEWS/D in 23 cases and inferior only in 7 cases. In the remaining 11 comparisons, no statistical inference could be reached. It is yet to be shown that the F-NEWS/D approach is to be preferred when using in the context of generalization rather than optimization.

TABLE VI: COMPARISON RESULTS – FRONT VS NO FRONT

Dataset	Obj.	F-NEWS/D	NF-NEWS/D	p-value
Wine	1	**9.7e-4(±1.2e-3)**	3.4e-3(±2.8e-3)	1.1e-6
	2	**2.0e-3(±1.3e-3)**	5.1e-3(±3.8e-3)	1.1e-6
	3	**3.5e-6(±1.6e-4)**	1.4e-4(±6.1e-4)	3.6e-3
UKM	1	**2.6e-3(±8.3e-8)**	3.8e-3(±3.4e-3)	4.3e-4
	2	**3.5e-2(±2.3e-3)**	4.0e-2(±7.3e-3)	8.1e-5
	3	**5.1e-2(±7.7e-3)**	5.4e-2(±6.2e-3)	1.6e-7
	4	**2.2e-3(±5.9e-4)**	4.4e-3(±9.5e-3)	2.4e-7
BT	1	1.1e-2(±5.1e-4)	1.1e-2(±8.8e-3)	<u>2.6e-1</u>
	2	2.0e-2(±1.6e-3)	2.0e-2(±4.6e-3)	<u>6.4e-1</u>
	3	**3.3e-2(±9.5e-4)**	3.4e-2(±2.0e-3)	3.7e-2
	4	1.9e-2(±7.5e-4)	1.9e-2(±4.3e-3)	1.0e-1
	5	5.0e-3(±7.0e-4)	**4.8e-3(±5.6e-3)**	1.1e-2
	6	1.0e-4(±1.6e-4)	**4.3e-5(±7.5e-3)**	1.5e-2
Zoo	1	**1e-36(±7.1e-26)**	4e-27(±4.0e-3)	5.4e-6
	2	**1e-26(±3.4e-20)**	3e-26(±4.9e-3)	1.9e-7
	3	**2.0e-3(±1.8e-3)**	2.5e-3(±3.1e-3)	2.7e-3
	4	1.8e-4(±8.7e-4)	6.2e-4(±6.4e-3)	<u>2.7e-1</u>
	5	2.3e-3(±2.7e-3)	7.5e-7(±3.6e-3)	<u>7.5e-1</u>
	6	**4e-18(±8.5e-15)**	1.6e-15(±4e-3)	1.8e-3
	7	2.5e-3(±5.6e-4)	2.5e-3(±4.8e-3)	<u>2.2e-1</u>

VII. SUMMARY AND CONCLUSIONS

This study suggested a modified decomposition framework for MaO-TWEANNs and an associated algorithm, which is termed NEWS/D. Eight benchmark problems are suggested based on existing multi-class classification datasets, ranging from three to seven objectives. Two types of evaluation methods are performed. First, an absolute evaluation is presented to check the algorithm's ability to reach the known ideal point. Second, two relative comparison techniques, based on multiple single-objective optimizations are proposed. The results, as obtained by the relative and absolute evaluations, demonstrate the applicability of the proposed algorithm. In addition, the results indicate that the objectives in the suggested benchmark problems are conflicting and that the obtained set of solutions includes more than one topology.

A numerical study towards a real-life application is also provided. The results of this additional study show that the development of class experts, for a multi-class classification ensemble might be supported by the proposed F-NEWS/D approach rather than the intuitive NF-NEWS/D approach. This could be explained by the idea of multi-objectivization [16].

Future research may focus on using the proposed framework to develop additional algorithms for MaO-TWEANNs. Such algorithms, including the current one, are expected to be examined on problems from areas such as robotics, control, and classification including the development of class experts. It should be re-noted that for classification, the algorithm should be revised to include generalization. As an alternative to the proposed decomposition framework, one might try to solve the problem using multi-factorial evolutionary algorithms [17].

ACKNOWLEDGMENT

The first author would like to acknowledge the generous scholarship that was provided by the Israeli Ministry of Science & Technology.

REFERENCES

[1] D. Floreano, P. Dürr, and C. Mattiussi, "Neuroevolution: From architectures to learning," *Evol. Intell.*, vol. 1, no. 1, pp. 47–62, 2008.

[2] K. O. Stanley, J. Clune, J. Lehman, and R. Miikkulainen, "Designing neural networks through neuroevolution," *Nat. Mach. Intell.*, vol. 1, no. 1, pp. 24–35, 2019.

[3] Y. Sun, B. Xue, M. Zhang, and G. G. Yen, "Evolving Deep Convolutional Neural Networks for Image Classification," *IEEE Trans. Evol. Comput.*, vol. 24, no. 2, pp. 394–407, 2020.

[4] O. Abramovich and A. Moshaiov, "Multi-objective topology and weight evolution of neuro-controllers," *in 2016 IEEE Congress on Evolutionary Computation, CEC 2016*, 2016, pp. 670–677.

[5] D. Nagar, A. Furman, and G. Nitschke, "The cost of complexity in robot bodies," *in 2019 IEEE Congress on Evolutionary Computation, CEC 2019 - Proceedings*, 2019, pp. 2713–2720.

[6] Showalter, I., & Schwartz, H. (2020). Objective Comparison and Selection in Mono- and Multi-Objective Evolutionary Neurocontrollers. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 2280–2287).

[7] K. O. Stanley and R. Miikkulainen, "Evolving neural networks through augmenting topologies," *Evol. Comput.*, vol. 10, no. 2, pp. 99–127, 2002.

[8] B. Li, J. Li, K. Tang, and X. Yao, "Many-objective evolutionary algorithms: A survey," ACM Comput. Surv., vol. 48, no. 1, p. 13, 2015.

[9] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Trans. Evol. Comput.*, vol. 11, no. 6, pp. 712–731, 2007.

[10] A. Trivedi, D. Srinivasan, K. Sanyal, and A. Ghosh, "A survey of multiobjective evolutionary algorithms based on decomposition," *IEEE Trans. Evol. Comput.*, vol. 21, no. 3, pp. 440–462, 2017.

[11] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, 2002.

[12] S. Künzel and S. Meyer-Nieberg, "Evolving artificial neural networks for multi-objective tasks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 10784 LNCS, pp. 671–686.

[13] A. Shenfield and S. Rostami, "Multi-objective evolution of artificial neural networks in multi-class medical diagnosis problems with class imbalance," in *2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2017*, 2017, pp. 1–8.

[14] Y. Tian, R. Cheng, X. Zhang, and Y. Jin, "PlatEMO: A MATLAB platform for evolutionary multi-objective optimization," *IEEE Comput. Intell. Mag.*, vol. 12, no. 4, pp. 73–87, 2017.

[15] M. Lichman, "UCI machine learning repository [http://archive.ics.uci.edu/ml].," *UCI Machine Learning Repository*. University of California, School of Information and Computer Science, 2013.

[16] Knowles, J. D., Watson, R. A., & Corne, D. W. (2001). *Reducing Local Optima in Single-Objective Problems by Multi-objectivization BT - Evolutionary Multi-Criterion Optimization* (E. Zitzler, L. Thiele, K. Deb, C. A. Coello Coello, & D. Corne (eds.); pp. 269–283). Springer Berlin Heidelberg.

[17] A. Gupta, Y. Ong, and L. Feng, "Multifactorial Evolution: Toward Evolutionary Multitasking," *IEEE Trans. Evol. Comput.*, vol. 20, no. 3, pp. 343–357, 2016.