Annals of Operations Research
https://doi.org/10.1007/s10479-022-05028-0

ORIGINAL RESEARCH

Enhancing the predictive performance of ensemble models through novel multi-objective strategies: evidence from credit risk and business model innovation survey data

Paritosh Jha[1] · **Marco Cucculelli**[2]

Accepted: 12 October 2022

Abstract
This paper proposes novel multi-objective optimization strategies to develop a weighted ensemble model. The comparison of the performance of the proposed strategies against simulated data suggests that the multi-objective strategy based on joint entropy is superior to other proposed strategies. For the application, generalization, and practical implications of the proposed approaches, we implemented the model on two real datasets related to the prediction of credit risk default and the adoption of the innovative business model by firms. The scope of this paper can be extended in ordering the solutions of the proposed multi-objective strategies and can be generalized for other similar predictive tasks.

Keywords Multi-objective optimization · Ensemble model · Prediction · Business model innovation · Credit risk

1 Introduction

Decision-makers in many areas, from industry to engineering and the social sector, consider multiple, conflicting objectives in their decision processes (Zhao, 2007). Standard statistical techniques for decision-making have mainly relied on different information criteria for modeling with single-objective functions (Burnham & Anderson, 2002). When several objectives have been considered, prevalent optimization techniques have aggregated multiple objective functions into single-objective functions for solving the problem, at the cost of excluding relevant alternatives or techniques that otherwise could be explored (Izui et al., 2015).

One promising approach for decision-making is the use of the ensemble model. The ensemble model combines multiple models using diverse available techniques and methods for enhancing predictive performance (Zhang & Yunqian, 2012). Despite a large volume of

✉ Paritosh Jha
paritoshjh1@gmail.com

Marco Cucculelli
m.cucculelli@staff.univpm.it

1 Department of Economics and Management, University of Bergamo, Bergamo, Italy

2 Department of Economics and Social Sciences, Marche Polytechnic University, Ancona, Italy

Published online: 07 November 2022

literature that advocates the benefits of the ensemble model based on single-objective function (Tumer & Ghosh, 1996), its limitations are well-known (Deb, 2001; Jin, 2006; Krawczyk, 2016; Wozniak et al., 2014).

By contrast, the multi-objective optimization approach provides an alternative tool that allows handling several conflicting objectives and constraints to identify optimal solutions. These solutions have been used for many different purposes, for example in the context of a regression problem (Breskvar et al., 2018), transfer learning (Kordík et al., 2018), meta-learning (Kordík et al., 2018), feature selection (Kou et al., 2021a, b; Kozodoi et al., 2019), (Ribeiro et al., 2020), the evaluation of different frameworks for driving ensemble model like competition and cooperation (Fletcher et al., 2020), or ensemble learning prediction-based strategies for re-initializing the sample of prediction (Sahâ et al., 2019; Shi et al., 2014; Wang, 2019; Wang et al., 2020), budget allocation problems (Kou et al., 2021a, b), for financial risk analysis (Kou et al., 2014; Li et al., 2021), bankruptcy prediction for SMEs (Kou et al., 2021a, b).

However, none of the studies in the literature considers the solution of the multi-objective optimization problem as a weight for combining a pool of candidate models to develop a weighted ensemble model. This paper addresses this topic by proposing four multi-objective approaches that have not only the potential to overcome the limitation of the single-objective optimization function, but also provide the advantage of integrating some prior knowledge in the estimation process to improve the predictive performance of the ensemble model.

Specifically, the paper provides a simple modeling framework and mathematical structure to solve the multi-objective problem to obtain local and global optimal solutions. The analytical solution assumes no bias and therefore ignores the problem that weights are random variate since weights are constrained to minimize any maximum deviations among the obtained solutions. Solutions of the multi-objective problems are then used as weights in the development of the ensemble model.

This paper provides three main contributions. First, it provides a set of four different multi-objective strategies that can be used to develop a weighted ensemble model from their analytical solutions as a linear combination of pooled models. This approach is new as the extant literature has addressed the issue in a very different way, specifically by doing feature selection and shuffling the sample for prediction. The pooled model can be parametric, non-parametric, or any general ensemble model. The combination process does not take into account any estimation of model parameters to avoid possible criticism summarized in Banner and Higgs (2017).

Second, it offers the flexibility to solve a multi-objective problem analytically without any need to reduce or transform the given constraints, thus avoiding the drawback of negatively affecting the performance of the predictive model when constraints and objectives are reduced. The solution obtained from the proposed multi-objective strategies is unordered and any ordering or preference is therefore not required in this scope of the study. Moreover, the proposed approach is rather simple, intuitive, and easy to implement compared to other multi-objective optimization techniques.

Third, it shows how the weighted ensemble model based on multi-objective approach helps to estimate class instances of supervised learning problems like credit risk default and business model innovation. The prediction of credit risk default and business model innovation as a supervised learning problem has not been widely explored in the literature from the perspective of multi-objective optimization to deal with conflicting objectives. This paper brings a novel direction in the definition of new perspectives to study the literature on credit risk and business model innovation. As for the generalization, the proposed approach can be applied to any problem that involves decision-making based on several constraints and conflicting objectives.

In this context, we compared the performance of a multi-objective strategy based on joint entropy against a single-objective optimization function on two real datasets. The performance comparison on the two real datasets suggests that the developed weighted ensemble model using a multi-objective strategy provides superior predictive performance compared to a single-objective optimization function. A similar comparison of the remaining proposed multi-objective strategies against single-objective function can be done, but we just preferred to make the comparison only picking the best of four proposed multi-objective strategies.

The proposed idea and insights from this paper are broad and can be an interesting to solve problems in other domains where multi-objective optimization function is of primary interest as a methodological approach and is not limited only to solve a supervised learning problem as discussed in this paper. For future research directions, any ordering of the solutions of multi-objective problems would be an additional advantage in improving the performance of the ensemble model.

The remaining section of the paper is as follows. Section 2 presents background information followed by the proposed strategies in Sect. 3 and results in Sect. 4. Section 5 presents the application of the theories followed by concluding remarks and future research direction in Sect. 6.

2 Background information

Any machine learning algorithm from the optimization point of view can be seen as a single objective learning, scalarized multi-objective learning, and Pareto-based Multi-objective learning. Single objective learning often minimizes mean squared error (MSE) on the training data but many different suitable error metrics can be equally used to solve optimization problem.

The recent trends in the use of machine learning algorithm has seen applications of multi-objective to overcome the limitations of single-objective functions depending on how cost function is adopted. The increasing impetus of multi-objective approach for various tasks is mostly attributed to the advancement of evolutionary algorithms and other stochastic search methods.

The advantage of multi-objective learning with scalar cost function helps in addressing different topics of machine learning such as clustering, feature selection, improvement of generalization ability, knowledge extraction, and ensemble model generation.

Using multi-objective optimization, one can categorize any learning problem as an optimization problem as it is often a task of model selection and parameter estimation evaluated against different criteria. For instance, in supervised learning, the common criteria is an error function that reflects the approximation quality whereas in the unsupervised problem, the criteria is to maximize inter-cluster similarity and minimize intra-cluster similarity. For problems of reinforcement learning, the criterion is a value function that helps in predicting the reward for an agent to perform a given action in a given state.

Many competing and conflicting objectives can be optimized together through the help of multi-objective optimization approach. For instance, complexity and interpretation of the model is one such conflicting objectives to be optimized together using multi-objective approach. These two objectives of the model are strongly interrelated to each other and in general, the lower is the complexity of the model, the easier it is to understand the model.

To obtain this, one has to consider a second objective reflecting the complexity of the model which can be aggregated as a scalar objective function keeping $f = E + \lambda\Omega$, where E is common error function, Ω is a measure for model complexity that indicates the number of free parameters in the model, with $\lambda > 0$ is a positive hyper-parameter.

Such approach is widely used to regularize neural networks, create interpretable fuzzy rules and many other usage in the field of machine learning. However, there are two main weaknesses in the use of scalarized objective function for multi-objective problem. Firstly, it is difficult to make an appropriate choice of hyper-parameter λ and secondly only single solution can be gained from which it is difficult to visualize any further additional insights into the problem. To overcome such limitation, one has to take advantage of a multi-objective approach that helps any learning algorithm to improve the overall accuracy.

3 Proposed strategies

Notation and assumptions

In this section, we propose four multi-objective strategies that can serve as a useful tool in enhancing the objectives of the weighted ensemble model. The models considered for weighted ensemble model are a collection of parametric, non-parametric, and ensemble model, to say, let $\{1, 2, \ldots, n\}$ represents the set of these models to which we assume allocating random vector of weights $w = \{w_1, w_2, \ldots, w_n\} \in \mathbb{R}^n$ as a preferential choice where w_i is any specific weight attached with any model $i \ \ for \ \ i = 1, 2, \ldots, n$ constrained as $\sum_{i=1}^{n} w_i = 1$.

Let $p = \{p_1, p_{2,\ldots,} p_n\} \in \mathbb{R}^n$ be the performance associated with each of the models. It may be possible that the weights may be mis-allocated due to existing co-variance. To minimize the co-variance between weights and performance of the models, we construct the following relation,

$$f_1(w) = w \cdot p^T = \sum_{i=1}^{n} w_i p_i \tag{3.1}$$

where $\sum(.,)$ is the co-variance between model and their allocated weights. Therefore, the error of the model is defined simply the co-variance between weights and transposed weights as the following,

$$f_2(w) = w \sum w^T \tag{3.2}$$

With the help of Eqs. (3.1) and (3.2), We can construct multi-objective optimization problem as bi-objective problem,

$$\min_{w \in C} \begin{bmatrix} w \ p^T \\ w \ \sum \ w^T \end{bmatrix}, \text{ where } C = \left\{ w \in \mathbb{R}^n; \sum_{i=1}^{n} w^T \ \mathbf{1} = 1, \ \ w^T \ \overline{\widehat{y}} = \overline{y} \ for \ i = 1, 2, \ldots, n \right\} \tag{3.3}$$

and $\mathbf{1}$ is a unitary matrix. The symbol $\overline{\widehat{y}}$, $\overline{y}$ refers to predicted and observed mean values.

Obtaining a solution for such minimization problem is difficult compared to single-objective problem. The scalarized approach discussed in Sect. 2 helps to certain extent for aggregating multi-objective function but is not efficient to obtain optimal solutions. To combat this limitation, one such approach is goal programming which is a special case of the multi-objective problem as a bi-objective problem where we fix a goal value for each objective function, and measure the deviations of the values of the objective function from their goal value over the feasible region.

The advantage of using such an approach is that we are able to optimize a target value for each goal function and then minimize the difference between each target function and its goal rather than directly optimizing goal objective function. Formally, the bi-objective problem can be reformulated as a goal programming problem by assigning to each f_i a goal value g_i and minimizing the deviation $(f_i - g_i)^+ \ for \ i = 1, 2$ over feasible region where + refers to the positive part of the function.

To be more precise, let us define a $g_1 = p*$,where $p*$ denotes the desired level of performance on the model and let $g_2 = 0$ as two target goal function. The goal vector $g = (p*, 0) \in\in \mathbb{R}^2$ do not necessarily lie in the objective space. Using the above settings, we propose four different multi-objective strategies that helps to reformulate multi-objective problem into bi-objective problem to enhance not only performance of the weighted ensemble model but also to trade-off against competing objectives of any given problem. The following flowchart explains the framework and its processing components for constructing weighted ensemble model from the proposed strategies.

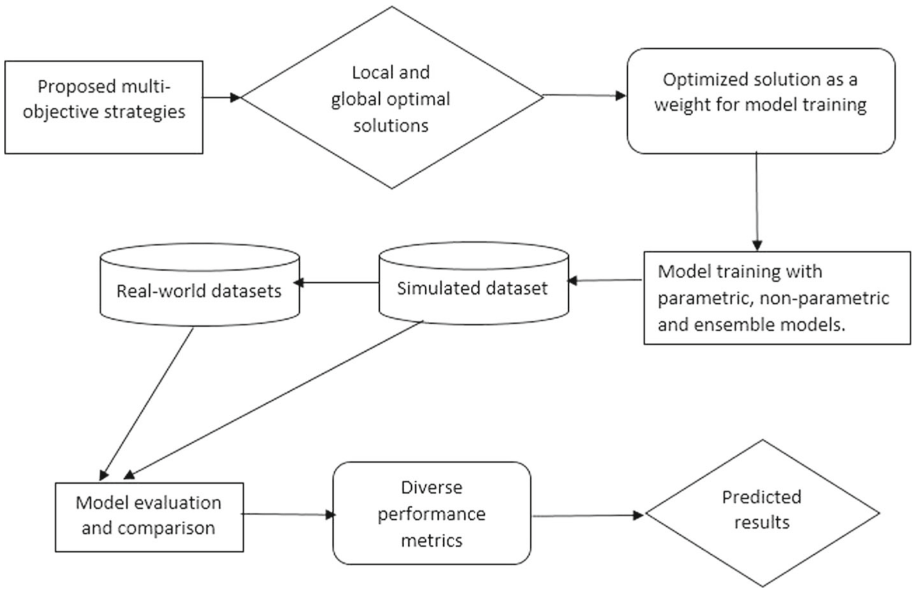

Fig. 1 Flowchart for weighted ensemble model from the proposed strategies

Strategy 1 The first strategy we consider is called weighted sum of deviations (WSD) that formally can be written as

$$\min_{w \in C} \sum_{j=1}^{N} w_j \ (f_j(w) - g_j)^+ \text{ where } C \text{ is a set of constraints.} \tag{3.4}$$

A scalarized or aggregated function can be written as $F = \sum_{j=1}^{N} \ w_j(f_j - g_j)^+$ which can further be formulated as convex combinations that can help us to generate a new curve as a weighted average of deviations for each objective from its goal.
Strategy 2 The second strategy we use here is called Chebyshev goal programming and can be considered as an extension of previous Strategy 1 since we try to minimize only the maximum weighted deviation instead of minimizing the sum of deviations. When this is done, this helps in minimizing other deviations which are smaller. More formally, we can write them as

$$\min_{x \in C} \left[\max_{j} \ w_j \ (f_j(x) - g_j)^+ \right] \ for \ j = 1, 2, \cdots N \text{ where C is any constraint defined in equation (3.3).} \tag{3.5}$$

Strategy 3 This strategy is called joint entropy and helps us to understand the uncertainty or divergence associated between models. The joint entropy of n models can be formulated as

$$H(x_1, \ldots, x_n) = - \sum_{x_1 \in \chi_1} \cdots \sum_{x_n \in \chi_n} P(x_1, \ldots, x_n) \ \log_2 [P(x_1, \ldots, x_n)] \tag{3.6}$$

More formally, to understand how much each of these models diverge from each other, we can formalize them as $x = (x_1, \ldots, x_n) \in p_n, \ y = (y_1, \ldots, y_n) \in q_n \ then \ for \ i =$ 2, ..., n , as it holds (3.7) $z_i = min \left\{ \sum_{j=1}^{i} p_j, \ \sum_{j=1}^{i} q_j \right\} - \sum_{j=1}^{i-1} z_j$ for any $z = x \wedge y$; $p_n \ and \ q_n$ which are respective marginal probability distributions. One has to keep in mind that such measure helps in understanding the diversification between models which is non-negative and concave.
Strategy 4: Another variant of Strategy 3 is to use cross entropy rather joint entropy for understanding diversification among models. The idea of using cross entropy is based on importance sampling. For instance, if we take a random sample $x_1, \ldots, x_n$ based on importance sampling with density g on χ and using unbiased estimator ℓ and likelihood ratio, we can minimize the distance of cross entropy which is equivalent to solving maximization problem

$$\max_{v} \int g^*(x) \ \ln \ f(x; v) \, dx \tag{3.8}$$

where $g^*(x) = \frac{I_{\{S(x) \geq \gamma\}} \ f(x;u)}{\ell}$ is the density measure and $f(.; v)$ is a family of density.

So far, we have been asserting that it is possible to formulate the given bi-objective problem into a goal programming problem to generate an optimal solution, but we do not know if the optimal solution obtained through goal programming problem is also the optimal solution to bi-objective problem. We can formalize a theorem in this context to see if it is true.

Theorem *If x^* is the optimal solution for goal programming then this also serves as a unique minimizer or optimal point for the multi-objective problem.*

Proof Using weighted sum of deviations method, we can approach to prove this theorem for the goal programming problem assuming that x^* is the unique global minimizer of

$$\min_{x \in C} \left[w_1 \left(xp^T - p^* \right)^+ + w_2 \left(x \sum x^T \right) \right] \tag{3.9}$$

Let us assume further that x^* is not a global optimal solution or optimal solution for the multi-objective problem, $\exists$ a point $\widehat{x} \in C$ with condition either $\widehat{x} p^T < x^* p^T$ *or* $\widehat{x} \sum \widehat{x} < x^* \sum x^{*^T}$. Therefore, we can say that following relation holds $w_1(\widehat{x}\ p^T - p^*)^+ + w_2(\widehat{x} \sum \widehat{x}) < w_1(x^* p^T - p^*)^+ + w_2(x^* \sum x^{*^T})$
where $\widehat{x}$ is a global minimizer of multi-objective problem and this is a contraction to what we assumed. An equivalent or alternate theorem can be established for global maximization problem to find optimal solutions of multi-objective problem. □

4 Results

Solving each of the strategies above using a minimization framework provides a set of non-dominated solutions that are not ordered but sufficiently serves as local and global optimal values for the considered objective functions. They are efficient solution which is used to rank the performance of machine learning models (parametric, non-parametric, and ensemble). This, in turn, helps us in mapping relationships between our objective function, which can be changed sequentially by varying weights especially in the strategic approach of a weighted sum of deviation, and Chebyshev goal programming.

The ordering analysis of the optimal solution is not considered in this scope of study in terms of *no preferred method, a priori method, posterior method, hybrid method, and interactive method* since these are broadly defined topics meeting the different purposes of solving a multi-objective problem. Our approach here to some extent is very similar to the no-preference method where we have been able to scalarize the problem taking the objectives that are normalized into a uniform dimensionless scale.

Each of the ensemble model based on proposed strategies were developed using parametric models (*logistic regression (GLM)*), and ensemble model average technique such as (*random forest, Bayesian moving average*).

To evaluate the performance of the ensemble model using proposed strategies on simulated data, various key performance metrics reflecting accuracy and error metrics were used such as hmeasure (H), Area under the roc curve (AUC), Minimum error rate (MER), and Minimum cost weighted error rate (MWL) which helps to examine predictive capability, discriminatory power and stability of the results. The simulated dataset is created randomly from the underlying structure of ensemble model development. The generated simulated dataset resembles real dataset as the data generating process is assumed to come from multi-variate normal distribution with similar mean and co-variances. The developed ensemble model is a weighted model of different classifiers which is combined using the optimal weight of proposed multi-objective problem.

Figure 2 presents the multi-objective solutions as an unordered point with two local minimum optimal solutions and a set of various points as a globally optimal solution with respect to Strategy 1 that is based on a weighted sum of deviations. The solution achieved through this strategy is useful for the direct comparison of objectives since unnecessary deviations are multiplied with weights to form a single sum for the goal or achievement function.

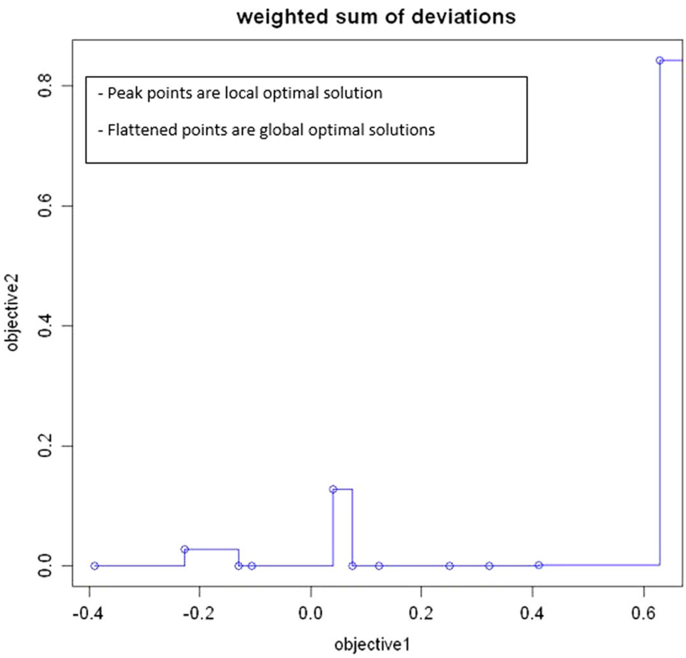

Fig. 2 Unordered solutions using Strategy 1

Figure 3 presents the multi-objective solutions with respect to Strategy 2 where we can see a peak at some point in their objective function value and being flattened at many other

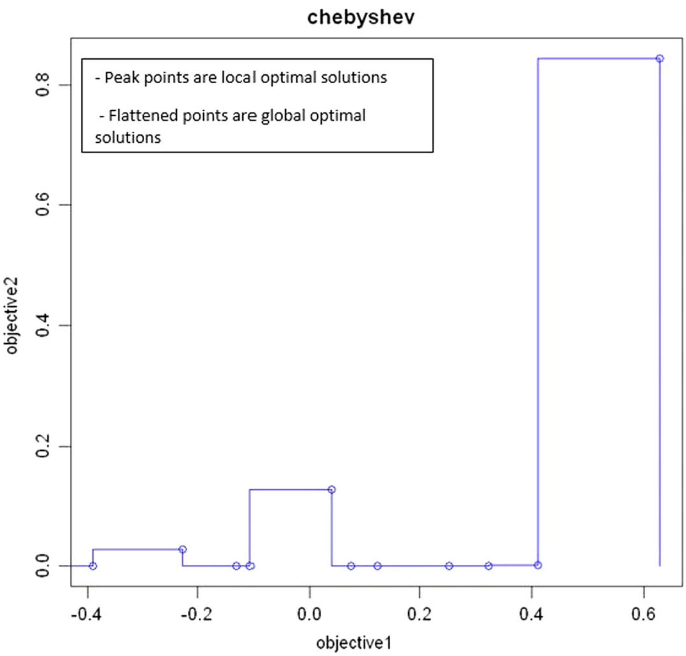

Fig. 3 Unordered solutions using Strategy 2

points with the goal to minimize maximum deviation in the goal programming approach i.e. to reduce maximum co-variance among the chosen machine learning models. These peak points are the local minimum optimal solution and flattened points are the global minimum optimal solution.

Figure 4 indicates the multi-objective unordered solution of the minimization problem referring to Strategy 3 that is based on joint entropy, and we can see multiple local and global optimal points.

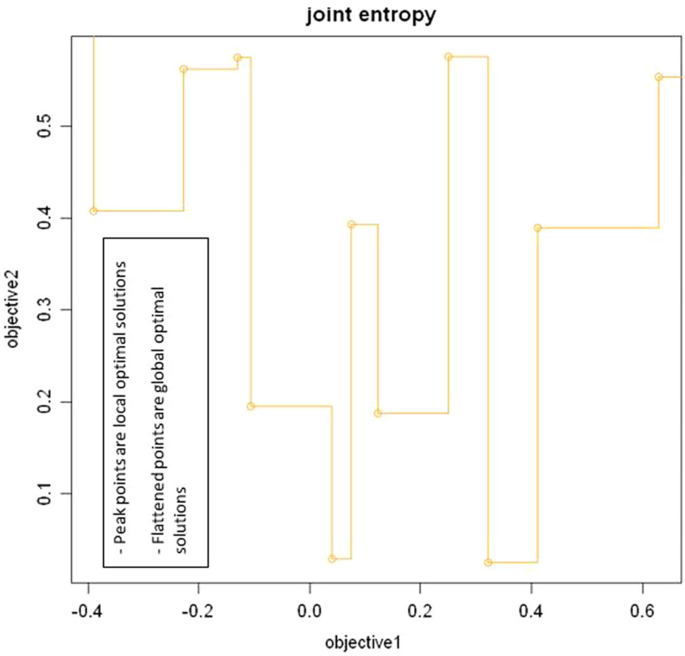

Fig. 4 Unordered solutions using Strategy 3

Figure 5 indicates the unordered solution of cross-entropy referring to Strategy 4 that shows multiple local optimal solutions and one global optimal solution.

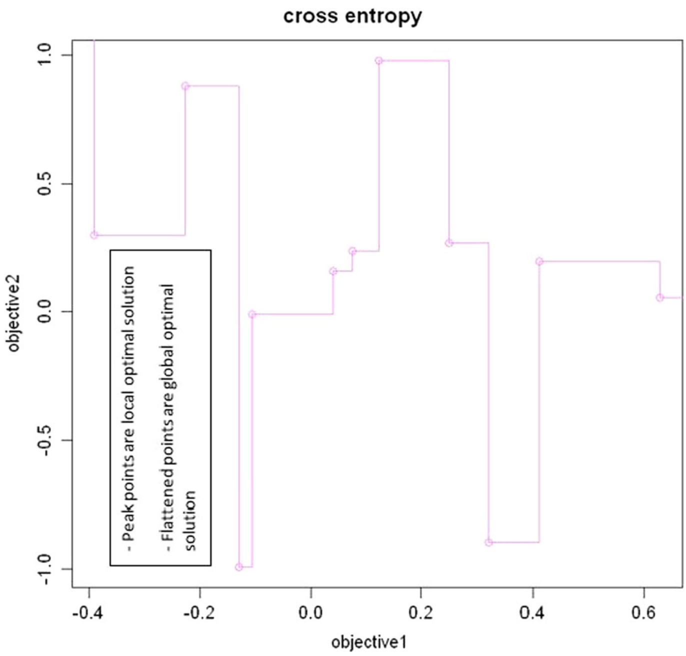

Fig. 5 Unordered solutions using Strategy 4

Table 1 Performance metrics reflecting accuracy of the strategies

Strategy	H	AUC
Strategy 1	0.60	0.92
Strategy 2	0.48	0.93
Strategy 3	**0.64**	**0.94**
Strategy 4	0.06	0.63

Significance of [bold]: The bold text is there to highlight best strategy among other strategies with respect to performance metrics

Table 2 Performance metrics reflecting error of the strategies

Strategy	MER	MWL
Strategy 1	0.14	0.14
Strategy 2	0.12	0.12
Strategy 3	**0.11**	**0.11**
Strategy 4	0.28	0.28

Significance of [bold]: The bold text is there to highlight best strategy among other strategies with respect to performance metrics

Tables 1, 2 indicate how each of the proposed strategy performs on simulated data with respect to the chosen performance metrics for developing ensemble model. The comparison in Tables 1, 2 suggests that proposed Strategy 3 based on joint entropy is superior to other proposed strategies that can help to enhance the predictive performance of ensemble model in classifying class instances of a classification problem.

5 Application

The applications of the proposed approach could be broad and various case studies of real-world problems can be solved using multi-objective optimization techniques. The application further depends as how many objectives have been defined and the level of interactions with decision makers. One can find various combinations of number of objectives and interaction to solve real-world problems from the multi-objective techniques (for instance, any optimization problem in policy planning and strategic management often requires various conflicting objectives to take into account for decision-making process). To facilitate the comprehension, we took a couple of dataset corresponding to different nature of problems and evaluated the application scope of the proposed approach.

The first real dataset for the application collects a credit risk information about the customer. The dataset has a binary dependent variable (ClientStatus) that takes class value 0 (good customer) and class value 1 (bad customer). A prior probability for the dependent variable shows 96.11 % of *class label 0* and 3.9 % of *class label 1*.This data is composed of 40,000 observations and 30 explanatory variables. The explanatory variable is mainly categorized as information about socio-demographic characteristics, customer equipment, customer history, and other things related to customer behavior. This dataset is provided by one of the leading financial institution in Europe and is not available as a public data for its wider use. Refer to appendix for more details on the dataset.

Another dataset used for the analysis from the application perspective of the proposed strategies is a survey data of Italian firms that report information on business model innovation summarized by major changes in the product, process, finance, and business network profiles of the firm. The dataset is composed of 7836 observations obtained from a questionnaire survey on a representative sample of Italian manufacturing firms, submitted in the period October 2019–March 2020. The explanatory variables mainly collect information on the economic and financial profile of the company, its innovation profile in terms of product, process and organizational innovations, the intensity of investment in IT-related assets, additional information on the ownership of the company and main individual characteristics of board members, managers and directors. The response variable is a binary choice variable indicating business model innovation with class *label 0* and class *label 1*. A prior probability for the dependent variable shows that 92.3 percent of *class label 0* indicate innovation in the business models and 7.7 percent of *class label 1* are firms that did not innovative their business model. Refer to appendix for more details on the dataset.[1]

It is obvious to see that the dependent variable for both the datasets is imbalanced class distribution and to train them for a predictive model, it is necessary to balance the class distribution using any over-sampling or under-sampling technique to avoid any over-fitting

[1] Business model innovation refers to changes in the existing structure of assets and operations (i.e., business model) that a company uses to deal with the market. Whether a firm at strategic level realizes this or not, business model is always there, which could be either in evident or latent form. In principle, the underlying logic or architecture of any business always refers to the business model in place. A firm changes many decisions at strategic level across various functions which may lead to the overall innovation of the existing business model. Specifically, it is the combination of several changes in different functions of business which matters the most in innovating the business model. The design of the survey is done in such a way to capture information of the changes done across different functions within the firm. Such integrated changes of different functions in the business most likely lead to innovation of the business model. In abstract sense, the binary dependent variable "business model innovation" is nothing but a function of individual indicators that refers changes in the business model. More precisely, business model innovation is a function BMI = f (f3, p7, f2, n5, ...), which is a linear combination of individual indicators of business model change. We thank an anonymous referee for helping to make this point clearer.

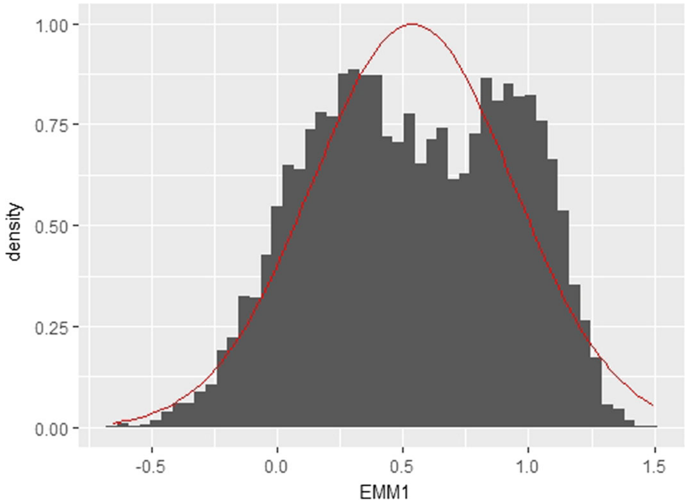

Fig. 6 Predicted distribution of ensemble model using multi-objective Strategy 3 on credit risk data

problem. To create class balance, we use SMOTE (synthetic minority over-sampling technique) which is a preferred technique to treat the imbalance problem of data. SMOTE (Chawla et al., 2002) creates synthetic observations based on existing minority observations that works on the principle of k-nearest neighbors. It generates new instances that are not just copies of the existing minority class: in fact, the rule is to take samples of feature space for each target class and its nearest neighbors. In this way, it increases the features available to each class and makes the samples more general and balanced.[2]

Although each of the four proposed strategies can be applied to any nature of datasets for a predictive model task, our focus is to pick the best proposed strategy and compare the performance against any single objective function that can be broadly defined as follows:

$$\min_{w} w^T \sum w \text{ Such that } w^T \mathbf{1} = 1 w^T \overline{\widehat{y}} = \overline{y} \tag{5.1}$$

From the four proposed multi-objective strategies, we compared Strategy 3 (joint entropy) against a single-objective function defined in Eq. 5.1. Figures 5, 6, 7 and 8 show the predicted probability distribution of Strategy 3 (joint entropy) and single-objective optimization on the credit risk data and business model survey data.

Looking at Table 3, Figs. 5 and 6, it infers that the performance of ensemble model using multi-objective optimization Strategy 3 based on joint entropy (EMM1) is superior in predicting credit risk default compared to single- objective optimization function (EMS1)[3]. The distribution of predicted probability in Fig. 5 is well separated for positive and negative predictive class in comparison to distribution in Fig. 6 which further confirms the superior performance of multi-objective optimization Strategy 3 based on joint entropy.

[2] For credit risk dataset, the class size, their distribution after SMOTE are (9342, 10,899) and (46%, 54%) respectively. For business model dataset, the class size, their distribution after SMOTE are (1881, 2037) and (48%, 52%) respectively.

[3] EMM1 refers the proposed strategy 3, EMM2 to strategy 4, EMM3 to strategy 2, and EMM4 to strategy 1. EMS1, EMS2, EMS3 and EMS4 refers to the single-objective optimization function of the proposed four strategies and follows the same sequence of EMM1, EMM2, EMM3, and EMM4.

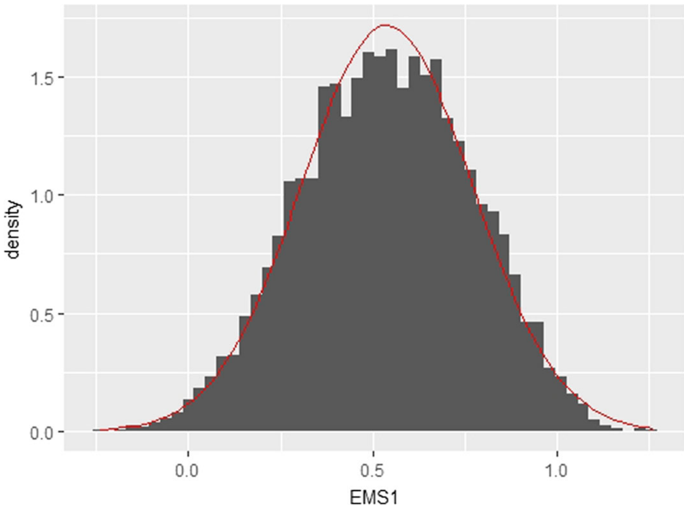

Fig. 7 Predicted distribution of ensemble model using single-objective function on credit risk data

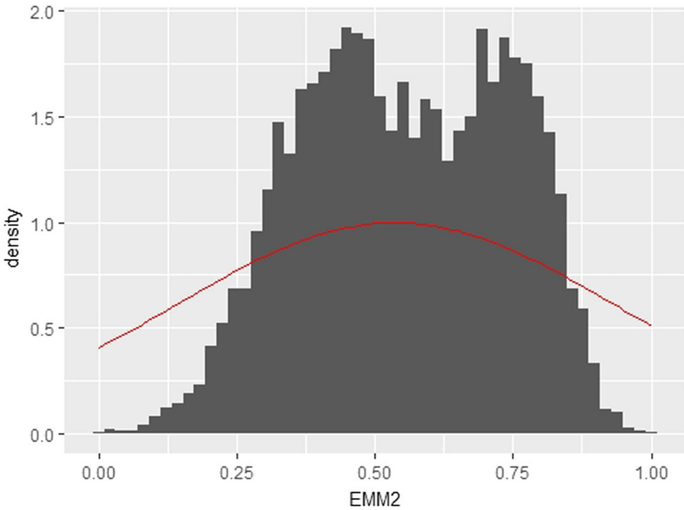

Fig. 8 Predicted distribution of ensemble model using Strategy 3 (multi-objective) on business model innovation survey data

Looking at Table 4 and Figs. 7, 8, it infers that the performance of ensemble model using multi-objective optimization Strategy 3 based on joint entropy (EMM) is superior in predicting business model innovation compare to its version of single-objective optimization function (EMS)[4]. The distribution of predicted probability in Fig. 7 is well separated for positive and negative predictive class in comparison to distribution in Fig. 8 which further

[4] EMM2 refers to the proposed multiobjective strategy 3 and EMS2 is a single-objective version of EMM2. EMM2 and EMM1 are interchangeably the same as they have been developed using strategy 3, it is just two different convention for evaluating the performance on two different datasets.So, is the case with EMS1 and EMS2. The other models in Table 4 stands for GLM(generalized linear model), RF(random forest), and BMA(Bayesian moving average).

Table 3 Performance evaluation of different ensemble model construction

Optimization function	H	AUC	MER	MWL
EMM1	0.64	0.94	0.11	0.11
EMM2	0.52	0.86	0.18	0.20
EMM3	0.66	0.89	0.16	0.18
EMM4	0.46	0.79	0.22	0.32
EMS1	0.05	0.63	0.25	0.27
EMS2	0.32	0.69	0.35	0.37
EMS3	0.03	0.64	0.25	0.22
EMS4	0.15	0.74	0.20	0.18

Table 4 Performance evaluation of multi-objective, single-objective ensemble model and other classifier

Optimization function	H	AUC	MER	MWL
EMM	0.56	0.95	0.17	0.12
EMS	0.45	0.76	0.16	0.18
GLM	0.26	0.64	0.36	0.48
RF	0.39	0.89	0.26	0.32
BMA	0.55	0.83	0.27	0.17

confirms the superior performance of multi-objective optimization Strategy 3 based on joint entropy.

To establish the statistical difference between obtained results, we used the DeLong test for comparing the difference in AUC values of the ensemble model based on single-objective optimization and multi-objective optimization function. The pairwise comparison of the model and using its p-value with reference to the significance value of 0.05, it is found that all pair is statistically significant and different except the pair of EMS1 and EMS2 in Table 3. In Table 4, it is found that all pair-wise comparison of models is statistically significant and different except for the pair RF and BMA (refer DeLong et al., 1988).

6 Conclusion

This paper gives a new perspective in connecting the use of the multi-objective approach to a classification problem by proposing different strategies to assess optimal solutions as a weight for developing ensemble models. This approach can help in understanding how to generate interpretable models, retrieve new insight for model selection, and model uncertainty. The overall benefit of the proposed strategies is to enhance the performance of the ensemble model for any predictive task that seeks the attention of multi-objective optimization.

Limitations to this approach mainly come from the difficulty to guarantee and measure convergence in order to achieve regular spacing of solutions, a result which is largely due to the dominance and diverse nature of multi-objective approaches. The research activity in the area of multi-objective optimization is an active field, with many challenging problems still open in the context of uncertainty handling, computational complexity, and robustness. For instance, one such intriguing question is the influences of learning behavior or simply a property of the learning curve due to a multi-objective approach to machine learning. Any

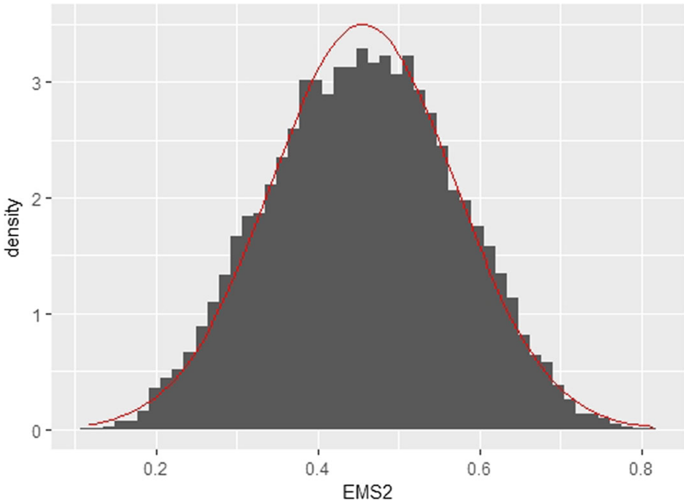

Fig. 9 Predicted distribution of ensemble model using single-objective function on business model innovation survey data

ordering of the optimal solutions obtained from the proposed strategies in this paper would be a useful direction to explore in future research.

7 Appendix

7.1 Details on credit risk dataset

See Tables 5, 6, 7 and 8.

Table 5 Socio-economic variable description

Variable	Description	Type
Age	Loan applicant age	Discrete
Region	Location details	Categorical
Account age	Age of the current account (expressed in years)	Discrete
Residency	Type of residence (owner or tenant)	Categorical
Residency age	Seniority of residence in the current residence (expressed in years)	Discrete
Civil status	Marital status (married, single, divorced …)	Categorical
Kids	Number of children	Discrete
Gender	Gender	Categorical
Income_applicant	Applicant income	Continuous
Income_family	Family income	Continuous
Profession	Job details	Categorical
Birth_place	Country of birth	Categorical
Job_seniority	Working seniority (expressed in years)	Discrete

Table 6 Client equipment variable description

Variable	Description	Type
Financing	Financing channel (agency, web, telephone …)	Categorical
Personal_loans	Current personal loans—number of practices	Discrete
Personal_balance	Current personal loans—residual amount on the balance	Continuous
Residual_duration	Current personal loans—residual duration to balance	Continuous
Approved_loans	Total finalized loans in progress—number of practices	Discrete
Approved_loans_balance	Total finalized loans in progress—remaining balance	Continuous
Approved_loans_maturity	Total finalized loans in progress—residual maturity at the balance	Continuous
Credit_card_info	Card—customer holding card	Discrete
Credit_card_balance	Card—credit card display	Continuous

Table 7 Client history variable description

Variable	Description	Type
Personal_loans_details	Personal loans paid in the last 24 months—number of files	Discrete
Approved_loans_details	Finalized loans paid in the last 24 months—number of practices	Discrete

Table 8 Client behavior variable description

Variable	Description	Type
Late_payment	Number of late payments from origin (in months)	Discrete
Behavioral_score_internal	Internal behavioral score	Continuous
Behavioral_score_bureau	Credit bureau behavioral score	Categorical
Recovery_yearly	Number of recovery ascents in the last 12 months	Discrete
Recovery_monthly	Number of months to recovery in the last 12 months	Discrete

7.2 Details on business model innovation dataset

See Table 9

Table 9 Variable description of business model innovation dataset

Variable	Description
f3	Did firm introduce new features to existing product to improve sales?
p7	Did firm make any change to products and process?
f7	Did firm set price dynamically?
p6	Did firm shorten the time to market?
f2	Did firm change the way in which they sell product (revenue model)?
n1	Did firm integrate their merger and acquisition upstream?

Table 9 continued

Variable	Description
n8	Did firm get grant to support innovation policy?
p3	Did firm introduce the product in niche market?
n3	Did firm integrate within the main business support activities?
n5	Did firm modify or introduce new direct sales channels (online, e-commerce,digital, and new sales network)?
f5	Did firm focus on differentiated product?
n2	Did firm integrate their merger and acquisition downstream?
o1	Did firm include any new function in their organizational process?
n7	Did firm sign any partnership with customer and supplier?
n3	Did firm integrate within the main business support activities?
p1	Did firm introduce new product?
n4	Did firm introduce business skill?
o7	Did firm change the hierarchy in organizational process?
f1	Did firm change the the pricing policy according to demand or discount system?
o2	Did firm delete any new function in their organizational process?
p5	Did firm introduce process innovation?
n6	Did firm modify or introduce new indirect sales channels (wholesalers, distributors and other intermediaries)?
p2	Did firm add additional service to existing product?
n5	Did firm introduce process innovation?
o4	Did firm introduce business skill?
p4	Did firm change the customer portfolio or market?
f6	Did firm use fixed price?
f4	Did firm focus on mass market?

References

Bäck, T. (1996). *Evolutionary algorithms in theory and practice: Evolution strategies, evolutionary programming, genetic algorithms*. Oxford University Press.

Banner, K. M., & Higgs, M. D. (2017). Considerations for assessing model averaging of regression coefficients. *Ecological Applications, 27,* 78–93.

Belton, V., & Stewart, T. (2002). *Multiple criteria decision analysis: An integrated approach*. Springer.

Breskvar, M., Kocev, D., & Džeroski, S. (2018). Ensembles for multi-target regression with random output selections. *Machine Learning, 107,* 1673–1709.

Burnham, K., & Anderson, D. R. (2002). Model selection and multi-model inference: A practical information-theoretic approach. *Springer, 26*(2), 1–488.

Chawla, Nitesh V., Bowyer, Kevin W., Hall, Lawrence O., & Philip Kegelmeyer, W. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research, 16,* 321–57.

Coello, C. A. C., Lamont, G. B., & Van Veldhuizen, D. A. (2007). *Evolutionary algorithms for solving multiobjective problems*. Springer.

Deb, K. (2001). *Multiobjective optimization using evolutionary algorithms* (pp. 1–518). Wiley.

Deb, K. (2001). *Multiobjective optimization using evolutionary algorithms*. Wiley.

Dellnitz, M., Schutze, O., & Hestermeyer, T. (2005). Covering pareto sets by multilevel subdivison techniques. *Journal of Optimization Theory and Applications, 124*(1), 113–136.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 837–845.

Ehrgott, M. (2005). *Multicriteria optimization*. Springer.
Ehrgott, M. (2012). Vilfredo Pareto and multiobjective optimization. Optimization stories. *Journal der Deutschen Mathematiker-Vereininggung, Extra, 21,* 447–453.
Fletcher, S., Verma, B., & Zhang, M. (2020). A non-specialized ensemble classifier using multi-objective optimization. *Neurocomputing, 409,* 93–102.
Ignizio, J. P. (1976). *Goal programming and extensions*. Lexington Books.
Izui, K., Yamada, T., Nishiwaki, S., & Tanaka, K. (2015). Multiobjective optimization using an aggregative gradient-based method. *Structural and Multidisciplinary Optimization, 51,* 173–182.
Jin, Y. (2006). *Multi-objective machine learning* (Vol. 14, pp. 1–660). Springer.
Jürgen, B., Kalyanmoy, D., Kaisa, M., & Roman, S. (2008). *Multiobjective optimization: Interactive and evolutionary approaches*. Springer.
Kordík, P., Černý, J., & Frýda, T. (2018). Discovering predictive ensembles for transfer learning and meta-learning. *Machine Learning, 107,* 177–207.
Kou, G., Peng, Y., & Wang, G. (2014). Evaluation of clustering algorithms for financial risk analysis using MCDM methods. *Information Sciences, 275,* 1–12.
Kou, G., Xu, Y., Peng, Y., Shen, F., Chen, Y., Chang, K., & Kou, S. (2021). Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection. *Decision Support Systems, 140,* 113429.
Kou, G., Xiao, H., Cao, M., & Lee, L. H. (2021). Optimal computing budget allocation for the vector evaluated genetic algorithm in multi-objective simulation optimization. *Automatica, 129,* 109599.
Kozodoi, N., Lessmann, S., Papakonstantinou, K., Gatsoulis, Y., & Baesens, B. (2019). A multi-objective approach for profit-driven feature selection in credit scoring. *Decision Support Systems, 120,* 106–117.
Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence, 5*(4), 221–232.
Kuncheva, L., & Whitaker, C. (2003). Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy". *Machine Learning, 51*(2), 181–207.
Li, T., Kou, G., Peng, Y., & Yu, P. S. (2021). An integrated cluster detection, optimization, and interpretation approach for financial data. *IEEE Transactions on Cybernetics*, 1–14.
Mackay, D. J. C. (2003). *Information theory, inferences, and learning algorithms*. Cambridge University Press.
Murphy, K. (2012). *Machine learning: A probabilistic perspective*. MIT Press.
Peimankar, A., Weddell, S. J., Jalal, T., & Lapthorn, A. C. (2018). Multi-objective en- semble forecasting with an application to power transformers. *Applied Soft Computing, 68,* 233–248.
Ribeiro, V. H. A., & Meza, G. R. (2020). Ensemble learning by means of a multi-objective optimization design approach for dealing with imbalanced data sets. *Expert Systems with Applications, 147,* 113–232.
Rosales-Perez, A., Garcia, S., Gonzalez, J. A., Coello, C. A. C., & Herrera, F. (2017). An evolutionary multi-objective model and instance selection for support vector machines with Pareto–based ensembles. *IEEE Transactions on Evolutionary Com- putation*, 1.
Sahâ, S., Sarkar, D., & Kramer, S. (2019). Exploring multi-objective optimization for multi-label classifier ensembles. *IEEE Congress on Evolutionary Computation (CEC), 2019,* 2753–2760.
Shi, C., Kong, X., Fu, D., Yu, P. S., & Wu, B. (2014). Multi-label classification based on multi-objective optimization. *Association for Computing Machinery, 5*(2), 1–22.
Smith, C., & Jin, Y. (2014). Evolutionary multi-objective generation of recurrent neu- ral network ensembles for time series prediction. *Neurocomputing, 143,* 302–311.
Tan, C. J., Lim, C. P., & Cheah, Y. N. (2014). A multi-objective evolutionary algorithm- based ensemble optimizer for feature selection and classification with neural network models. *Neurocomputing, 125,* 217–228.
Tumer, K., & Ghosh, J. (1996). Analysis of decision boundaries in linearly combined neural classifiers. *Pattern Recognition, 29*(2), 341–348.
Wang, W., et al. (2019). An effective ensemble framework for multiobjective optimization. *IEEE Transactions on Evolutionary Computation, 23*(4), 645–659.
Wang, F., Li, Y., Liao, F., & Yan, H. (2020). An ensemble learning based prediction strategy for dynamic multi-objective optimization. *Applied Soft Computing, 96,* 106592.
Wozniak, M., Graña, M., & Corchado, E. (2014). A survey of multiple classifier systems as hybrid systems. *Information Fusion, 16,* 3–17.
Zhang, C., & Yunqian, M. (2012). *Ensemble machine learning: Methods and applications*. Springer.
Zhao, J., Jiao, L., Xia, S., Basto Fernandes, V., Yevseyeva, I., Zhou, Y., Emmerich, T. M., & M. (2018). Multi-objective sparse ensemble learning by means of evolutionary algorithms. *Decision Support Systems,111*, 86–100. https://doi.org/10.1016/j.dss.2018.05.003
Zhao, H. (2007). A multi-objective genetic programming approach to developing Pareto optimal decision trees. *Decision Support Systems, 43*(3), 809–826.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.