Poster

TITLE

Optimizing the Efficiency / Adverse Impact Trade-off in Personnel Classification Decisions

ABSTRACT

The paper presents an analytic method for estimating the efficiency and the adverse impact of general personnel classification decisions. Additionally, the method is integrated in a decision making framework to obtain predictor composites that show Pareto-optimal efficiency / adverse impact trade-offs in a mixture population classification context.

PRESS PARAGRAPH

Different subgroups display different means on specific performance predictors, leading to the quality- diversity dilemma in the personnel selection context. However, since classification situations still arise in practice, the reality of effect sizes will lead to adverse impact in these personnel decision situations as well. The current method to estimate the classification efficiency given a set of predictors, different subgroups and their characteristics, was extended to yield the adverse impact ratio as well. Additionally, this method was implemented in an algorithm that leads to predictor weights that result in optimal trade-offs between efficiency and diversity.

Authors' Note: Correspondence regarding this article should be sent to Celina Druart. Email: Celina.Druart@UGent.be

This paper was presented at the 25th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, Ga: April 2010.

The term "personnel decisions" refers to all situations that require deciding about how to make the best use of available talent, with selection and classification decisions as typical instances. Compared to selection, classification decisions have received much less interest in the research literature during the past decades. For example, the prevailing attention for the quality-diversity dilemma in the personnel selection literature has not yet extended to classification decisions. However, classification situations arise in practice, and implementing a selection instead of a classification perspective in these situations leads to erroneous expectations concerning the efficiency of the personnel decision. Moreover, practitioners are presently without guidance in designing classification decisions that aim to balance the efficiency and the adverse impact of intended classification decisions as applied to applicant groups that are a mixture of different subpopulations.

Consequently, this paper aims at extending the current analytic method to estimate the classification efficiency (De Corte, 2000) to the case where applicants come from several subpopulations. Additionally, following the example of De Corte, Lievens, and Sackett (2007), who combined predictors to achieve optimal trade-offs between selection quality and adverse impact, the extended method will be integrated in a multi-objective optimization framework so as to achieve optimal trade offs between efficiency and diversity in a classification context.

The following section recapitulates the basic concepts as well as the relevance of the classification perspective. Next, the available analytic methods for estimating the classification efficiency are resumed. Then the extension of the currently most general method is presented and it is shown how the method can be integrated in a decision making framework to obtain a summary of the Pareto-optimal efficiency/adverse impact trade-offs that can be achieved for general classification decisions. Finally, we present an application that illustrates the potential of the new method.

The Classification Perspective

Classification decisions select from one applicant pool for different jobs/ positions simultaneously, thus evaluating in which trajectory a certain individual would be expected to achieve more than in another. Classification therefore represents an extension of selection in that it considers multiple positions at the same time, and does not necessarily reject a precentage of the applicants (Alley, 1994). Personnel and educational decisions that favour adopting a classification perspective arise in settings that require assessing a group of individuals in the light of several different open positions. Possible applications can be found in large industrial or governmental organizations, the armed forces as well as educational settings.

Adopting a classification instead of a selection perspective when appropriate, can substantially increase the efficiency of high stakes personnel decisions. Brogden (1951) illustrated this fact already, by comparing the outcomes of assignment to different positions using a single predictor versus seperate differential predictors. The increased efficiency of classification compared to selection is to be ascribed to a more favourable selection ratio per criterion, as explained by Brogden (1951).

Classification requires constructing different predictor composite scores from an available test battery, one for each of the C different jobs, for all applicants. This implies that differential validity of the predictors is a prerequisite to perform classification. The main arguments against the viability of differential prediction relate to the controversy between the general mental ability and the specific aptitude theory for predicting and/ or explaining educational or job performance (Zeidner & Johnson, 1994). Yet, despite the evidence that specific aptitudes/ composites show little if any incremental validity over the general mental ability factor (Viswesvaran & Ones, 2002), it does not follow that differential prediction and classification should be abandoned. As pointed out by Zeidner & Jonhson (1994), the main issue is not whether composites exhibit incremental validity over and above general mental ability, but rather whether composites show differential validity across different jobs. Hunter and Hunter (1984) already found evidence that the validity of GMA as a predictor decreases from .56 to .23, for the highest and lowest complexity jobs. And not only cognitive performance predictors give rise to differential validity; validity levels of personality measures are susceptible to moderation too (e.g., Barrick & Mount, 1991). Thus, we follow Zeidner & Johnson (1994) and Bobko (1994) in their conclusion that the differential validity hypothesis is sustained and as a consequence, a classification approach remains a viable option.

The primordial goal of a classification decision (and other personnel decisions) is to maximize the *classification efficiency* which corresponds to the expected *actual* criterion score of the accepted and classified individuals. To achieve this purpose, candidates are optimally assigned on the basis of their *estimated* criterion scores which typically correspond to regression based predictor composite scores, one for each criterion. In general, *optimal assignment* encompasses the optimization of two decisions: an accept/reject decision for each individual, and for each accepted applicant, a decision on the job she/ he will be assigned to.

Analytic estimation of classification efficiency

An important notion concerning the determination of the classification efficiency is the *allocation average*, also known as the *mean predicted performance*, or *MPP* (Scholarios, Johnson, & Zeidner, 1994). This concept was first defined by Brogden (1955) as the expected *estimated* criterion score of the assigned applicants. However, Brogden (1955) proved that the allocation average equals the average *actual* performance score when the assignment is based on regression weighted composite scores. Together with the obvious fact that regression weighted composites are the best possible predictors of future criterion performance, this explains why all analytic methods for estimating the classification efficiency focus on situations where regression weighted predictor composites govern the assignment process.

All presently available methods to estimate the classification efficiency also depend on the optimal assignment procedure as described by Brogden (1954). Brogden (1954) proved that optimal assignment can be obtained by first finding a set of appropriate constants, $\mathbf{k} = (k_1, \ldots, k_C)'$, which are added to the criterion estimates, $\mathbf{E} = (E_1, \ldots, E_C)'$, of all applicants, in order to obtain so-called *augmented* criterion estimates, $\mathbf{V} = \mathbf{E} + \mathbf{k} = (V_1, \ldots, V_C)'$ and fill the required quota per criterion. Optimal assignment is then achieved by assigning individuals to the position for which they have the highest augmented criterion estimate.

De Corte (2000) developed the thusfar most general analytic method for calculating the classification efficiency. Assuming that the available predictors follow a joint multivariate distribution, he derives the density function of the highest augmented criterion estimate, $f(v^{(h)})$ with $v^{(h)}$ the value of the highest augmented criterion estimate $V^{(h)}$ (i.e., $V^{(h)} = \max(V_1, \ldots, V_C)$) and the corresponding, criterion-specific relative densities, $f_c(v^{(h)})$. Subsequently, these relative density functions are used to solve a system of nonlinear equations, so as to determine the augmentation constants. Finally, based on Brogden's (1955) equality, he derives the expected criterion score of assigned individuals as the expectation over the appropriate relative density of the highest augmented estimated criterion value, minus the product of the quota and the constants summed over all criteria.

Efficiency and Adverse Impact of General Classification Decisions

The quality-diversity dilemma has thusfar only been considered in the context of selection decisions, but it is obvious that, given applicants coming from different subgroups with different mean values on the predictor variables, classification decisions will suffer from adverse impact as well. Therefore, it is important to not only focus on the expected efficiency, but take into account the work force diversity outcome too. Additionally, studies on selection decisions demonstrated that other than regression based composites can lead to a better balance between quality and diversity than the balance achieved by regression weighted composites (De Corte et al., 2007; De Corte, Lievens, & Sackett, 2008). As a consequence, it seems essential to develop a method for estimating the double outcome of classification efficiency and adverse impact when assignment is based on general, compared to regression weighted, predictor composites. This development depends on certain additional assumptions and requires addressing three issues: conceiving general estimated criteria, deriving the density function of the highest augmented estimated criterion score in a mixture population context and computing the conditional expected criterion given the value of the highest augmented (non-regression based) estimated criterion score. These assumptions and issues are briefly discussed next.

For the present mixture population context it is assumed that in the G different subpopulations, criteria $\mathbf{Y}$ have the same variance and predictors $\mathbf{X}$ have the same validities, variances and covariances, but differ in their mean values. As in De Corte (2000) the unit variance predictors have a multivariate normal distribution in each subpopulation g, $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}_{\mathbf{X}g}, \mathbf{R}_{\mathbf{X}})$. Also, all criterion variances are equated to one, and $\boldsymbol{\mu}_{\mathbf{Y}g} = (\mu_{Y_1 g}, \ldots, \mu_{Y_C g})'$ denotes the group-specifc mean criterion vector. For each criterion c, we make the slightly more general assumption that $(Y_c, \mathbf{X}')$ have a multinormal distribution.

Corresponding to what is the case for regression based criterion estimates, we propose using general criterion estimates with variance equal to the squared validity of the predictor composite. In particular, we consider criterion estimates E_c equal to $E_c = \frac{\rho_{Z_c Y_c} Z_c}{\sigma_{Z_c}}$, with $Z_c = \mathbf{w}'_c \mathbf{X}$ the corresponding predictor composite and $\mathbf{w}_c$ the arbitrary non-negative weights of the predictors in the composite.

In the mixture context the applicant pool consists of several different subgroups, with different mean values on predictors and criteria. As a consequence, the density function of the highest augmented criterion estimate also becomes a mixture of different density functions, one for each subgroup, with mixture proportions, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_G)'$, equal to the proportional representation of the respective subgroups in the total applicant pool. The equations in the system of non-linear equations that solves for the classification constants are adapted accordingly.

When the assignment of applicants to trajectories is based on general criterion estimates, the value of the conditional expected criterion score of an assigned candidate from group g is no longer a simple function of the corresponding augmented criterion estimate. However, using a result from Waldman (1984), it can be shown that the value can be equated to the mean of a properly defined conditional truncated multinormal distribution which can be computed using Tallis (1961) formula on the expectation of variables that follow a truncated normal distribution.

Using some earlier introduced notation and $E_g(Y_c|V^{(h)} = v^{(h)} = V_c)$ to represent the conditional expected criterion score of an individual from group g with value $v^{(h)}$ for the highest augmented criterion estimate occuring for criterion c, the above developments then imply that the expected criterion score of an individual from group g that is assigned to this criterion, $E_g(Y_c^{(a)})$ can be equated to

$$E_g(Y_c^{(a)}) = E_{f_{cg}}(E_g(Y_c|V^{(h)} = v^{(h)} = V_c))$$

where $E_{f_{cg}}$ indicates expectation over the relative density f_{cg} of $V^{(h)}$ occuring for criterion c in group g. Next, noting that the expected criterion score of individuals assigned to criterion c, $E(Y_c^{(a)})$, equals $\sum_g \tau_g E_g(Y_c^{(a)})$, the efficiency of the classification, B, can be computed as $B = \frac{\sum_c q_c E(Y_c^{(a)})}{\sum_c q_c}$, where q_c denotes the required classification proportion for job c.

The extent of adverse impact can be gauged using any of a number of different adverse impact indicators. In line with the selection literature, we will focus on the commonly used adverse impact ratio (AIR). Using the previously introduced (relative) density functions f_{cg}, the proportion of a specific subpopulation that is assigned to a specific criterion as well as the total classification proportion, over all criteria, for a specific subgroup are easily obtained. From these proportions, the adverse impact ratio can then be calculated as the ratio of the total classification rate in a specific minority subgroup and the total classification rate in the majority subgroup.

Obtaining Pareto-optimal Trade-offs between Classification Efficiency and AI

Given the classification parameter data (e.g., the make up of the applicant population, the predictor characteristics and their relation to the jobs) that characterize the classification scenario, using different sets of criterion estimates (i.e., different sets of predictor composites) will lead to different trade-offs in terms of classification efficiency and AI. The issue then becomes that of finding the sets of criterion estimates that lead to efficiency/AI trade-offs that can not be bettered by any other set of criterion estimates. The latter trade-offs are typically referred to as Pareto-optimal and any set of criterion estimates that corresponds to a Pareto-optimal trade-off is called Pareto-optimal as well.

Observe that a similar issue has been treated before by De Corte, Lievens & Sackett (2007, 2008), albeit in a selection, instead of the present classification decision context. To resolve the issue, these authors propose adopting a multi-objective optimization framework. Because this framework represents a generic approach to solving for Pareto-optimal solutions in general multi-objective decision contexts, we decided to follow the same strategy for obtaining Pareto-optimal trade-offs in the present classification context. However, because of the numerical complexities involved in calculating the classification efficiency and AI for given sets of criterion estimates, we adopted an evolutionary multi-objective optimization program instead of the more classical normal boundary intersection variant (Das & Dennis, 1998) implemented by De Corte et al. (2007). More specifically, we approximate the set of Pareto-optimal classification efficiency and AI trade-offs using the nondominated sorting genetic algorithm II (NSGA-II) of Deb, Pratap, Agarwal, & Meyarivan, 2002.

Implementation and illustration

We wrote a computer program that integrates the present analytic method to obtain the outcomes of general classification decisions within the NSGA-II routine. To illustrate the method and its implementation in the genetic algorithm, data are borrowed from a study by Johnson, Abrahams, and Held (2004), reporting intercorrelations and effect sizes of 9 subtests from the ASVAB (Armed Services Vocational Aptitude Battery), and their validities regarding 32 Navy occupations. Throughout the example, it is assumed that the total applicant pool consists of 78% majority (White) en 22% minority (Black) applicants. Three of the 9 subtests were selected as predictors: Verbal (VE), Automotive-Shop Information (AS) and Coding Speed (CS). Their intercorrelations and effect sizes are shown in Table 1, while their validities concerning 5 Navy occupations are displayed in Table 2. The 5 positions are Aviation Boatswain's Mate - Fuels (ABF), Aviation Structural Mechanic - Equipment (AME), Construction Mechanic (CM), Aircrew Survival Equipmentman (PR) and Steelworker (SW).

Given the aforementioned values of the classification parameters, the proposed procedure was used to analyze the optimal adverse impact - classification efficiency trade-offs for an intended classification system with an overall selection rate of 50%. The criterion specific selection rates are 8, 11, 9, 12 and 10%. Because the scenario encompasses three predictors to optimally assign individuals to one of the five criteria, each complete set of decision variables consists of 15 predictor weights, which are restricted between 0 and 1. After 25 generations (i.e., iterations), the NSGA-II algorithm resulted in the Pareto-optimal front, this is a collection of Pareto-optimal efficiency/AI trade-offs, plotted in Figure 1. Table 3 provides additional details on a selected number of the obtained optimal trade-offs. These selected trade-offs are indexed by a red diamond in Figure 1, and for each of them, Table 3 indicates the value of the classification efficiency and the adverse impact ratio objective, as well as the value of the decision variables that correspond to the Pareto-optimal trade-off.

– insert Figure 1 about here –

For example, using Figure 1 and Table 3, it can be verified that Pareto-optimal trade-off point 3 corresponds to an AIR of .69 and a classification efficiency of .34. To properly interpret the latter value, it must be compared to the expected job performance of randomly classified individuals which equals $-.15$ for the present application parameter data. Also, because trade-off number 3 is Pareto-optimal, no other set of predictor weights (and, hence, no other set of composites and estimated criteria) will do at least as well on one of the objectives and, at the same time, do better on the other objective.

As a further example, consider Pareto-optimal trade-off point 1 in Figure 1, representing an AIR of .31 and a classification efficiency of .48. These outcomes correspond to using regression weighted criterion estimates, and represent the optimal classification efficiency. All other Pareto-optimal trade-offs result in a lower efficiency, but at the same time in a higher AIR. On the other end of the Pareto front, we find Pareto-optimal trade-off point 5, with an efficiency of .13, and the highest attainable level of AIR in this scenario: .86. All other Pareto-optimal trade-offs result in a lower AIR, but at the same time in a higher efficiency.

Discussion

Previous research on personnel decisions has shown that adopting a classification instead of a selection perspective in an appropriate context can lead to an increased level of expected job performance of the assigned individuals. However, given the reality of effect sizes of certain predictors, adverse impact is expected to arise in classification situations just as it does in selection situations so that the quality of the assigned applicants may no longer be the single focus of classification decisions. In response to the already in the selection literature vigorously debated quality/adverse impact issue, the paper is the first to present an analytic method to estimate the efficiency as well as the adverse impact of classification decisions. In addition, the new method is implemented in a genetic algorithm in order to obtain the predictor weights that lead to Pareto optimal trade-offs between the goals of efficiency and diversity. These Pareto-optimal trade-offs represent the optimal levels of AIR obtainable for each level of classification efficiency, as well as the optimal levels of efficiency, for all attainable levels of AIR.

In the absence of further information, there is no single Pareto-optimal trade-off that can be said to be better than the other trade-offs. Decisions about the relative value of the Pareto-optimal solutions depends on the amount of efficiency one is willing to give up, in order to reach a lower level of adverse impact. It is clear though, that general, compared to regression weighted, predictor composites lead to much more balanced efficiency-diversity trade-offs. Also, as shown in the example application, the method permits a better informed design of composite predictors to perform classification decisions for which efficiency as well as diversity are important goals.

Apart from its advantages, the present method has certain limitations as well. In particular, its results are based on certain assumptions about the distribution of predictors and criteria in the different subgroups, but identical assumptions have been invoked by previous studies of the effects of predictor weighing on the balance between selection quality and diversity (De Corte et al., 2007) and on classification efficiency (Scholarios et al., 1994). Also, some of these assumptions such as, for example, the assumption that all criteria have equal variance can be relaxed to extend the method to the case where some positions are more critical to fill than others.

In summary, practitioners and researchers interested in the outcome of classification decisions, based on specific predictors and their characteristics, are provided with a method that yields the expected efficiency and the adverse impact ratio of these planned classification decisions. In addition, they can dispose of a decision aid in designing predictor composites that offer a Pareto-optimal balance between the goals of quality and work force diversity. Finally, as illustrated in the example, using general instead of regression weighted predictor composites leads to a much wider range of attainable diversity levels.

References

Alley, W. E. (1994). Recent advances in classification theory and practice. In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification: New directions* (pp. 431–442). Hillsdale, NJ: Erlbaum.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1–26.

Bobko, P. (1994). Issues in operational selection and classification systems: Comments and commonalities. In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification: New directions* (pp. 443–456). Hillsdale, NJ: Erlbaum.

Brogden, H. E. (1951). Increased efficiency of selection resulting from replacement of a single predictor with several differential predictors. *Educational and Psychological Measurement*, *11*, 173–196.

Brogden, H. E. (1954). A simple proof of a personnel classification theorem. *Psychometrika*, *19*, 205–208.

Brogden, H. E. (1955). Least squares estimates and optimal classification. *Psychometrika*, *20*, 249–252.

Das, I., & Dennis, J. E. (1998). Normal-boundary intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems. *SIAM J. Optimiz.*, *8*, 631–657.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, *6*, 182–197.

De Corte, W. (2000). Estimating the classification efficiency of a test battery. *Educational and Psychological Measurement*, *60*, 73–85.

De Corte, W., Lievens, F., & Sackett, P. (2007). Combining predictors to achieve optimal trade-offs between selection quality and adverse impact. *Journal of Applied Psychology*, *92*, 1380–1393.

De Corte, W., Lievens, F., & Sackett, P. (2008). Validity and adverse impact potential of predictor composite formation. *International Journal of Selection and Assessment*, *16*, 183–194.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72–98.

Johnson, J. W., Abrahams, N., & Held, J. D. (2004). *A procedure for selecting predictors considering validity and adverse impact.* Poster presented at the 19th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago.

Scholarios, D. M., Johnson, C. D., & Zeidner, J. (1994). Selecting predictors for maximizing the classification efficiency of a battery. *Journal of Applied Psychology*, *79*, 412–424.

Tallis, G. M. (1961). The moment-generating function of the truncated multinormal distribution. *Journal of the Royal Statistical Society, Series B*, *23*, 223–229.

Viswesvaran, C., & Ones, D. S. (2002). Agreements and disagreements on the role of general mental ability (gma) in industrial, work, and organizational psychology. *Human Performance*, *15*, 211–231.

Waldman, D. M. (1984). The mean of the conditional, truncated multinormal distribution. *Communications in Statistics - Theory and Methods*, *13*, 2679–2689.

Zeidner, J., & Johnson, C. D. (1994). Is personnel classification a concept whose time has passed? In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification: New directions* (pp. 377–410). Hillsdale, NJ: Erlbaum.

Table 1

The matrix $\boldsymbol{R}_{\mathbf{X}}$ of predictor correlations, and their effect sizes

	Predictor intercorrelations			Effect sizes
Predictors	VE	AS	CS	White - Black d
VE	1.000			0.684
AS	0.506	1.000		1.213
CS	0.337	0.029	1.000	0.178

Note. The three predictor variables are Verbal (VE), Automotive- Shop Information (AS) and Coding Speed (CS) ASVAB subtests.

Table 2

The matrix $\boldsymbol{D}$ of predictor validities

	Criteria				
Predictors	ABF	AME	CM	PR	SW
VE	0.469	0.344	0.446	0.457	0.499
AS	0.524	0.404	0.546	0.487	0.646
CS	0.257	0.231	0.238	0.393	0.176

Note. Predictors are Verbal (VE), Automotive- Shop Information (AS) and Coding Speed (CS) ASVAB subtests. The criteria are five Navy occupations, namely Aviation Boatswain's Mate - Fuels (ABF), Aviation Structural Mechanic - Equipment (AME), Construction Mechanic (CM), Aircrew Survival Equipmentman (PR) and Steelworker (SW).

Table 3

Selected pareto-optimal classification efficiency/ adverse impact trade-offs

			Predictor weights		
Optimal trade-off	Adverse impact	Classification efficiency	VE	AS	CS
1	.3142	.4812	0.1966	0.4193	0.1786
			0.1076	0.3442	0.1848
			0.1532	0.4635	0.1729
			0.1381	0.4074	0.3347
			0.1897	0.5472	0.0962
2	.5000	.4315	0.8603	0.6983	0.5635
			0.3545	0.0422	0.7160
			0.2483	0.4989	0.1727
			0.1667	0.2271	0.6259
			0.3111	0.6164	0.0261
3	.6905	.3406	0.4477	0.0770	0.8550
			0.0094	0.0016	0.6827
			0.0669	0.2042	0.6373
			0.0271	0.1407	0.7818
			0.4757	0.9879	0.0671
4	.7967	.2515	0.2158	0.0643	0.9848
			0.0043	0.0027	0.9276
			0.0135	0.0775	0.6491
			0.0668	0.0000	0.6304
			0.4644	0.6290	0.8227
5	.8551	.1311	0.1132	0.1213	0.8110
			0.0132	0.0152	0.7463
			0.0374	0.0086	0.3587
			0.0377	0.0260	0.9512
			0.0052	0.0005	0.9080

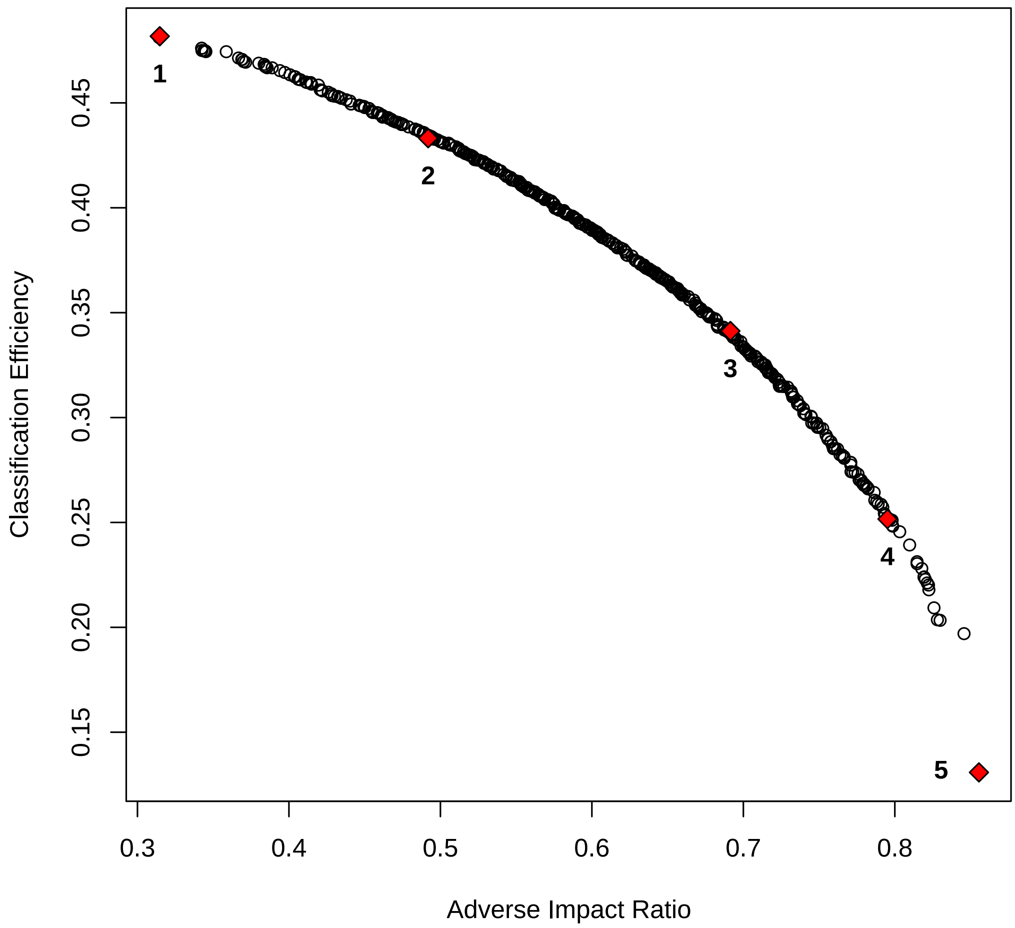

Figure 1. Pareto surface with Pareto-optimal efficiency/ AI trade-offs