ESENSC: A Polynomial-Time Axiomatic Alternative to SHAP

Kazuhiro Hiraki
International Monetary Fund
khiraki@imf.org

Shinichi Ishihara
Independent Researcher
ishihara5683@gmail.com

Takumi Kongo*
Fukuoka University
kongo@adm.fukuoka-u.ac.jp

Junnosuke Shino†
Waseda University
junnosuke.shino@waseda.jp

Abstract

We propose ESENSC, a computationally efficient and theoretically grounded alternative to SHAP. ESENSC is a deterministic, closed-form attribution rule that requires only linearly many coalition-value evaluations, satisfies the null-player property, and admits an axiomatic characterization based on efficiency, a restricted form of differential marginality, and explicit computational constraints. Empirically, on tabular prediction tasks using neural network and XGBoost models, ESENSC shows close agreement with exact SHAP in both attribution magnitude and feature rankings while requiring substantially less computation. Under the evaluated settings, it compares favorably with widely used sampling-based SHAP approximations. While exact SHAP requires exponentially many coalition evaluations, ESENSC scales linearly with the number of features. These results show that ESENSC offers a favorable trade-off between agreement with exact SHAP and computational cost, providing a practical and theoretically grounded alternative for feature attribution in high-dimensional applications.

1 Introduction

Additive feature attribution (AFA) methods decompose a model prediction into feature-level contributions and play a central role in explainable artificial intelligence (XAI). Among them, SHAP, based on the Shapley value from cooperative game theory [13], has emerged as the de facto standard due to its strong theoretical guarantees and model-agnostic applicability [10]. Formally, given a trained model f and an observation x_k, an AFA allocates the difference between the prediction $f(x_k)$ and the baseline $E[f(X)]$ across features according to their contributions. SHAP can be interpreted as the Shapley value of a cooperative game in which features are treated as players. We refer to this formulation as a XAI–TU game.

In this paper, we develop alternative attribution rules based on two classical solution concepts from cooperative game theory: equal surplus (ES) and proportional allocation (PA), and systematically compare their properties with those of SHAP. There are two main motivations for seeking such alternatives. First, while SHAP is theoretically elegant, its exact computation requires evaluating an exponential number of coalitions (2^n), creating a formidable barrier in high-dimensional settings where the number of features is large. In contrast, ES and PA rules are well known in cooperative game theory as solution concepts with low computational complexity. Second, existing approaches to addressing this computational challenge — namely, approximation methods for SHAP — often

*https://sites.google.com/site/takumikongo/
†https://sites.google.com/site/junnosukeshino/

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

lack a clear theoretical foundation and may exhibit instability or limited accuracy in practice. This suggests the need for attribution methods that are both computationally efficient and theoretically well grounded.

For ES–type solutions, we construct a novel attribution rule, ESENSC, based on an equal weight combination of Equal Surplus and Egalitarian Nonseparable Contribution (ENSC) [3, 4]. While the naive combination is computationally attractive, it violates the null-player property. Such violations are undesirable because features with no effect on predictions should receive zero attribution. We address this issue by modifying the rule to enforce the null-player property, resulting in ESENSC, which admits a closed-form representation and can be computed in polynomial time.

For PA–type solutions, we first address the order-reversal problem, which arises when positive and negative coalition values coexist. In such cases, proportional allocation can reverse the ranking of feature contributions. We develop modified PA–type AFAs, called PARPA, that preserve the ordering of features even under sign conflicts.

We evaluate the proposed methods on tabular prediction tasks using neural networks and gradient boosting models (with additional results for random forests in the appendix). The results show that ESENSC achieves substantial computational gains over exact SHAP while closely matching its attribution patterns, often outperforming widely used SHAP approximation methods in both deviation and rank agreement. In contrast, PA–type methods exhibit instability and large deviations, consistent with their theoretical properties in XAI–TU games.

Furthermore, we provide theoretical foundations by establishing an axiomatization of ESENSC. We show that, for $n \neq 3$, this attribution rule is uniquely determined by efficiency, the null-player property, restricted differential marginality, pairwise interaction splitting and information parsimony. The substantive case is $n \geq 4$: for $n \leq 2$, every game is 2-additive and ESENSC coincides with SHAP, whereas for $n = 3$ SHAP also satisfies all five axioms. This characterization shows that ESENSC can retain several SHAP–compatible axiomatic properties while achieving computational savings under an explicit linear information constraint. In this context, the solution concept most closely related to ESENSC is FESP proposed by [2]; however, FESP is primarily heuristic and lacks a full axiomatic characterization.

ESENSC is particularly effective in high-dimensional settings where exact SHAP is computationally infeasible and sampling-based approximations may exhibit variability in their estimates, offering a scalable and theoretically grounded alternative. To the best of our knowledge, ESENSC is the first model-agnostic AFA that provides a unique axiomatic characterization for $n \geq 4$ under explicit computational constraints, thereby addressing a fundamental trade-off between interpretability, computational efficiency, and theoretical rigor.

In summary, this paper makes three main contributions. First, we formalize feature attribution problems as XAI–TU games and highlight structural properties specific to this setting. Second, we propose ESENSC, a computationally efficient AFA that achieves strong empirical agreement with SHAP while substantially reducing computational cost. Third, we establish an axiomatic characterization of ESENSC, clarifying how computational tractability can be achieved while preserving key desirable properties. Together, these results position ESENSC as a principled and practical alternative to SHAP for feature attribution.

The remainder of the paper is organized as follows. Section 2 introduces XAI–TU games. Section 3 develops AFA methods. Section 4 presents empirical results. Section 5 provides the axiomatic characterization. Section 6 concludes.

2 Explainable AI TU Games (XAI–TU games)

Let t denote the number of observations and n the number of features. Define the index sets $N = \{1, \ldots, n\}$ and $T = \{1, \ldots, t\}$. Let $X = (X_1, \ldots, X_j, \ldots, X_n)$ be the $t \times n$ feature matrix. For each feature $j \in N$, let $X_j = (x_{1,j}, \ldots, x_{t,j})'$ denote the vector of its realizations across observations. For each observation $\tau \in T$, let $x_\tau = (x_{\tau,1}, \ldots, x_{\tau,j}, \ldots, x_{\tau,n})$ denote the feature vector of the τ-th observation. Let f be a trained prediction model, and let $Y = (y_1, \ldots, y_t)'$ denote the vector of predicted values given by f, i.e., $Y = f(X)$.

For any coalition $S \in 2^N$, define $x_{\tau,S} = \{x_{\tau,j} \mid j \in S\}$, that is, the subvector of the τ-th observation consisting of the features in S. Similarly, define $X_S = \{X_j \mid j \in S\}$. Let $s = |S|$ denote the

cardinality of S. Whenever no confusion arises, a singleton coalition $\{i\}$ is simply written as i. In cooperative game theory, a transferable-utility (TU) game is represented by a pair (N, v), where $N = \{1, \ldots, n\}$ is the set of players and v is a real-valued function defined on 2^N. A central problem in cooperative game theory is how to allocate the total worth $v(N)$ among players, given the worth $v(S)$ that each coalition $S \subseteq N$ can generate on its own.

We now construct, for each observation τ, a TU game in which the set of features N is the set of players. For the τ-th observation, define a characteristic function $v_\tau : 2^N \to \mathbb{R}$ by $v_\tau(S) = \mathbb{E}\left[f(x_{\tau,S}, X_{N\setminus S})\right]$, where the expectation is taken with respect to a reference distribution of the features in $N \setminus S$. The TU game defined by this expression is referred to as a XAI–TU game.

Throughout this paper, we adopt the interventional formulation of the characteristic function. Specifically, for a coalition S, the feature values in $N \setminus S$ are drawn jointly from their empirical distribution, independently of $x_{\tau,S}$. Accordingly, $v_\tau(S)$ is defined as $\mathbb{E}[f(x_{\tau,S}, X_{N\setminus S})] = \frac{1}{t}\sum_{\tau' \in T} f(x_{\tau,S}, x_{\tau',N\setminus S})$, which corresponds to the interventional version of SHAP commonly used in the literature. We do not consider conditional expectations of the form $\mathbb{E}[f(X) \mid X_S = x_{\tau,S}]$.

$v_\tau(\emptyset) = \mathbb{E}[f(X)]$ represents the baseline prediction when all features are treated as unknown. Unlike standard cooperative game formulations where $v(\emptyset) = 0$ is often assumed without loss of generality, in XAI–TU games $v_\tau(\emptyset)$ is generally nonzero and must be treated explicitly.

Another notable feature of the XAI–TU games is that the values $v_\tau(S)$ may differ in sign across coalitions. The total surplus to be allocated, namely the improvement of the prediction from its baseline level, $v_\tau(N) - v_\tau(\emptyset)$, may be either positive or negative. Similarly, marginal contributions such as $v_\tau(i) - v_\tau(\emptyset)$ may differ in sign across features (players). It is even possible for $\sum_{i\in N}\big(v_\tau(i) - v_\tau(\emptyset)\big)$ to have a different sign from $v_\tau(N) - v_\tau(\emptyset)$. Such coexistence of positive and negative contributions is not typical in traditional cooperative game theory but arises naturally in XAI–TU games. When standard solution concepts, typically proportional allocation rules, are applied directly, this may lead to unintuitive attribution patterns. These issues are examined in the next section.

For later analysis, we introduce Harsanyi dividends [5, 13] and the class of 2-additive games. For a game v, let $\lambda^v_\emptyset = v(\emptyset)$ and, for every nonempty $T \subseteq N$, $\lambda^v_T = v(T) - \sum_{R\subsetneq T} \lambda^v_R$; equivalently, $\lambda^v_T = \sum_{R\subseteq T}(-1)^{|T|-|R|}v(R)$, so that $v(S) = \sum_{R\subseteq S}\lambda^v_R$ for all $S \subseteq N$. Since $v(\emptyset)$ is generally nonzero in XAI–TU games, it is absorbed into $\lambda^v_\emptyset$; in particular, $\lambda^v_{\{i\}} = v(i) - v(\emptyset)$. We write λ^v_i for $\lambda^v_{\{i\}}$. A game v is 2-additive if $\lambda^v_T = 0$ for every $T \subseteq N$ with $|T| \geq 3$.

Finally, in many XAI applications the number of features n is large. Since the exact computation of the Shapley value grows exponentially in n, computing exact SHAP values becomes computationally infeasible in most practical settings. This motivates the search for computationally efficient alternatives.

These characteristics distinguish XAI–TU games from standard cooperative games and motivate the development of new solution concepts tailored to this setting.

3 Solution Concepts for XAI–TU Games

3.1 Additive Feature Attribution (AFA)

In a XAI–TU game, a solution concept for a given observation τ allocates the difference $v_\tau(N) - v_\tau(\emptyset)$ across features according to their contributions. We refer to any such allocation rule as an additive feature attribution (AFA). In what follows, we suppress the index τ and write the associated TU game as (N, v). For each feature $j \in N$, let $\psi_j \in \mathbb{R}$ denote its assigned contribution, and write $\psi = (\psi_1, \ldots, \psi_n)$. We say that ψ is an AFA if

$$\sum_{j\in N} \psi_j = v(N) - v(\emptyset).$$

That is, an AFA exactly distributes the difference between the full-information prediction and the baseline prediction. In the remainder of this section, we compare AFAs derived from cooperative game–theoretic solution concepts.

3.2 SHAP

SHAP in a XAI–TU game is defined as

$$\psi_j^{\mathrm{SHAP}} = \sum_{S \subseteq N \setminus j} \frac{s!(n-s-1)!}{n!} \left(v(S \cup j) - v(S)\right).$$

This coincides with the Shapley value of the associated XAI–TU game [13, 10]. SHAP has a strong theoretical foundation as the Shapley value, but its exact computation requires evaluating $v(S)$ for all 2^n coalitions, resulting in exponential complexity.

In cooperative game theory, equal surplus (ES) and proportional allocation (PA) rules are well known for their low computational cost. Building on these tractable solution concepts, we formulate corresponding AFAs and systematically analyze their properties in XAI–TU games.

3.3 AFAs Based on Equal Surplus and Related Solutions

The equal surplus solution (ES) is defined by

$$\psi_j^{\mathrm{ES}} = v(j) - v(\emptyset) + \frac{v(N) - v(\emptyset) - \sum_{i \in N} \{v(i) - v(\emptyset)\}}{n}.$$

The ES allocation proceeds in two steps. First, each player j receives its marginal contribution relative to the empty coalition, $v(j) - v(\emptyset)$. Second, the residual surplus, defined as $v(N) - v(\emptyset) - \sum_{i \in N}(v(i) - v(\emptyset))$, is distributed equally among all players. The ES–type AFA ψ_j^{ES} was briefly noted in [2, 6]. Here, we further develop ES–type AFAs for XAI–TU games.

The first related AFA can be viewed as a "reverse" ES rule, known as the Egalitarian Nonseparable Contribution (ENSC), formalized and axiomatized in cooperative game theory ([3, 4, 11, 12]). While ES builds from the empty coalition by assigning marginal contributions $v(j) - v(\emptyset)$, ENSC starts from the grand coalition and assigns the marginal loss $v(N) - v(N \setminus j)$ to each player j. As in ES, the remaining surplus is then equally redistributed among all players.

Formally, the ENSC-based AFA is defined as

$$\psi_j^{\mathrm{ENSC}} = v(N) - v(N \setminus j) + \frac{v(N) - v(\emptyset) - \sum_{i \in N} \{v(N) - v(N \setminus i)\}}{n}.$$

A further AFA can be obtained by taking the simple average of ES and ENSC. To facilitate later modifications, we explicitly refer to this unadjusted version as ESENSC-unadj:

$$\psi_j^{\mathrm{ESENSC\text{-}unadj}} = \frac{\psi_j^{\mathrm{ES}} + \psi_j^{\mathrm{ENSC}}}{2}.$$

A key advantage of ES–type solutions is their low computational cost. While SHAP requires evaluating $v(S)$ for all 2^n coalitions, ES and ENSC each require only $n+2$ evaluations, and ESENSC-unadj requires $2n + 2$, which is still substantially smaller than 2^n. As shown in the experiments, the differences in computation time among ES, ENSC, and ESENSC-unadj are negligible. Accordingly, we use ESENSC-unadj as a preliminary benchmark within the class of ES–type AFAs.

While ESENSC-unadj is computationally efficient, it has a key drawback: it typically violates the null-player property.

- **Null–player property.** If $j \in N$ satisfies $v(S \cup j) = v(S)$ for all $S \subseteq N \setminus j$, then $\psi_j(v) = 0$.[3]

This violation arises from the unconditional equal redistribution of residual surplus, which assigns non–zero payoffs to players with zero marginal contributions. To address this issue, we consider modified ESENSC rules that restrict redistribution to features with non–zero marginal effects.

While multiple modifications are possible, we focus on a parsimonious version that satisfies the null-player property:[4]

[3] We call player j satisfying this condition a null player in v, and occasionally write ψ_j as a function of its associated (XAI-)TU game v.

[4] An alternative modification proceeds as follows: first assign $v(j) - v(\emptyset)$ and redistribute the residual surplus equally among players with $v(j) \neq v(\emptyset)$; then assign $v(N) - v(N \setminus j)$ and redistribute the residual among players with $v(N) \neq$

- Step 1: Assign $v(j) - v(\emptyset)$ to each player j.
- Step 2: Assign $v(N) - v(N \setminus j)$ to each player j.
- Step 3: Take the average of these two quantities.
- Step 4: Redistribute the remaining surplus equally among players whose values are nonzero in at least one of Steps 1 or 2.

Formally, the modified rule, called ESENSC, admits the following closed-form expression, which is equivalent to the above procedure. First, for each $j \in N$, define

$$b_j(v) = \frac{v(j) - v(\emptyset) + v(N) - v(N \setminus j)}{2}.$$

Second, let $\mathcal{A}(v)$ denote the set of players $k \in N$ for which $v(k) \neq v(\emptyset)$ or $v(N) \neq v(N \setminus k)$. We call a player in $\mathcal{A}(v)$ boundary–active and a player outside $\mathcal{A}(v)$ boundary–inactive. Moreover, let $\mathbb{V} = \{\, v : 2^N \to \mathbb{R} \mid \mathcal{A}(v) \neq \emptyset \,\}$. Hereafter, we restrict attention to XAI–TU games in $\mathbb{V}$. ESENSC is defined for games in $\mathbb{V}$.

Lastly, define the residual required for efficiency by

$$R(v) = v(N) - v(\emptyset) - \sum_{i \in \mathcal{A}(v)} b_i(v).$$

Since $b_i(v) = 0$ for every $i \notin \mathcal{A}(v)$, this is equivalent to subtracting $\sum_{i \in N} b_i(v)$ from the total surplus.

With these definitions, ESENSC assigns $\psi_j^{\mathrm{ESENSC}} = 0$ to every $j \notin \mathcal{A}(v)$, and for $j \in \mathcal{A}(v)$,

$$\psi_j^{\mathrm{ESENSC}} = b_j(v) + \frac{R(v)}{|\mathcal{A}(v)|}. \tag{1}$$

This modified AFA satisfies the null-player property. In fact, ESENSC satisfies a stronger condition, possibly assigning zero to some players who are not null players. This condition is not imposed for normative reasons. Rather, under computational constraints, it allows features that may behave as null players (though they are not necessarily true null players) to be assigned zero attribution. In the subsequent analysis, we adopt ESENSC as the representative ES–type AFA.[5]

The equal residual redistribution in ESENSC also admits a simple optimization interpretation. Conditional on assigning zero attribution to players outside $\mathcal{A}(v)$, ESENSC is the unique efficient allocation that minimizes the squared Euclidean adjustment from the preliminary boundary scores $b_i(v)$. Thus, equal residual redistribution is the minimum-Euclidean correction required to restore efficiency. Appendix A provides a formal statement and proof.

It is also useful to relate ESENSC directly to SHAP through Harsanyi dividends introduced in Section 2, which decompose a game into coalition–specific value components. Appendix B shows that ESENSC coincides with exact SHAP for every 2–additive game. More generally, when all players are boundary-active, any ESENSC–SHAP difference is generated by Harsanyi-dividend components associated with coalitions of size three or larger. Appendix B provides the exact decomposition, a corresponding bound, and a characterization of the boundary-inactive case.

3.4 AFAs Based on Proportional Solutions and Related Solutions

We next consider AFAs based on proportional allocation (PA).[6] The PA rule is defined as:

$$\psi_j^{\mathrm{PA}} = \frac{v(j) - v(\emptyset)}{\sum_{i \in N} \{v(i) - v(\emptyset)\}} \cdot \{v(N) - v(\emptyset)\}. \tag{2}$$

$v(N \setminus j)$; finally, average the two outcomes. This rule is well defined if (i) $\exists j$ such that $v(j) \neq v(\emptyset)$ and (ii) $\exists j$ such that $v(N) \neq v(N \setminus j)$; otherwise it is undefined. In practice, these conditions rarely fail, so the rule behaves similarly to ESENSC. However, from a theoretical perspective, ESENSC is defined on a strictly larger class of games.

[5] ESENSC is not continuous in general because the boundary–active set $\mathcal{A}(v)$, and hence the set over which the residual is redistributed, may change discontinuously as v varies. On regions where $\mathcal{A}(v)$ remains unchanged, ESENSC is continuous.

[6] The solutions discussed in this subsection are defined only for games in which the denominator is not zero.

Furthermore, as in Subsection 3.3, one may also consider a "reverse" proportional rule defined by the following ([14, 7, 8, 18]):

$$\psi_j^{\mathrm{ROP}} = \frac{v(N) - v(N \setminus j)}{\sum_{i \in N} \{v(N) - v(N \setminus i)\}} \cdot \{v(N) - v(\emptyset)\}.$$

and the simple average of PA and ROP:

$$\psi_j^{\mathrm{PAROP}} = \frac{\psi_j^{\mathrm{PA}} + \psi_j^{\mathrm{ROP}}}{2}.$$

Those PA–type rules satisfy the null-player property but face a fundamental difficulty in XAI–TU games. To illustrate, consider PA (same arguments hold for ROP and PAROP). As noted in Section 2, $\sum_{i \in N}(v(i) - v(\emptyset))$ and $v(N) - v(\emptyset)$ may have opposite signs. In such cases, $v(i) - v(\emptyset) < v(j) - v(\emptyset)$ with $v(i) - v(\emptyset) \neq 0$ and $v(j) - v(\emptyset) \neq 0$ may imply $\psi_i^{\mathrm{PA}} > \psi_j^{\mathrm{PA}}$, reversing the ordering of marginal contributions.[7] We refer to this as the order-reversal problem, illustrated in the following example.

Example 1. *Consider the two-player game given by $v(\emptyset) = 0$, $v(1) = -3$, $v(2) = 2$, and $v(\{1, 2\}) = 1$. Then $\psi_1^{\mathrm{PA}} = 3$ and $\psi_2^{\mathrm{PA}} = -2$. Feature 1 decreases the prediction, whereas feature 2 increases it. Yet PA assigns a large positive contribution to feature 1 and a large negative one to feature 2.*

If an AFA does not induce the order-reversal problem, we say that it satisfies the order preservation property. As a PA–type AFA which satisfies this property, we define the following PARPA. Appendix C presents the derivation of PARPA and explains why it avoids this problem:

$$\psi_j^{\mathrm{PARPA}} = \begin{cases} \psi_j^{\mathrm{PA}} & \text{if } \left[v(N) - v(\emptyset)\right] \cdot \left[\sum_{i \in N} \{v(i) - v(\emptyset)\}\right] > 0, \\ \psi_j^{\mathrm{RPA}} & \text{otherwise,} \end{cases}$$

where

$$\psi_j^{\mathrm{RPA}} = \frac{v(N) - v(\emptyset) - \sum_{i \neq j} \{v(i) - v(\emptyset)\}}{\sum_{i \in N} \left\{v(N) - v(\emptyset) - \sum_{k \neq i} \{v(k) - v(\emptyset)\}\right\}} \cdot \{v(N) - v(\emptyset)\}$$

4 Experiments

We evaluate the AFA methods introduced in Section 3 by comparing deviation from exact SHAP and computational cost. Experiments in this section are based on the California Housing dataset, whereas supplementary results using the HELOC dataset are presented in Appendix D.

Dataset. We use the California Housing dataset,[8] a widely used benchmark for regression tasks, with median house value as the target variable. To study scalability, we augment the original eight features with additional synthetic features drawn from a standard normal distribution and standardize all inputs. The number of features is increased up to 512, allowing us to examine both approximation accuracy and computational scaling as the dimensionality grows.

Models. We consider two widely used nonlinear regression models: XGBoost and a feedforward neural network. For both models, hyperparameters are tuned via cross-validation over standard ranges, following common practice in tabular prediction tasks. For more details, see Appendix E.

Attribution Methods. We compare three categories: (i) exact SHAP as the benchmark, (ii) our AFA methods from Section 3 (including ESENSC and PARPA as representative ES–type and PA–type rules, along with related variants), and (iii) widely used SHAP approximations, namely Kernel SHAP ([10]), Permutation SHAP ([15, 16]),[9] and TreeSHAP ([9]). For TreeSHAP, we employ two different settings: (i) TreeSHAP (auto), which uses the default conditional distribution without background data, and (ii) TreeSHAP (interv.), which utilizes background data for interventional expectations.

[7] We exclude potential null players (i.e., i with $v(i) - v(\emptyset) = 0$) from the order comparison. Under PA, such players receive zero attribution; moreover, efficiency, the null-player property, and order preservation extended to potential null players are incompatible in general.

[8] Available in `scikit-learn` via the `fetch_california_housing` function (1990 U.S. Census data).

[9] For Permutation SHAP, the number of model evaluations is fixed at $(2n + 1) \times 10$ to ensure comparability across feature dimensions.

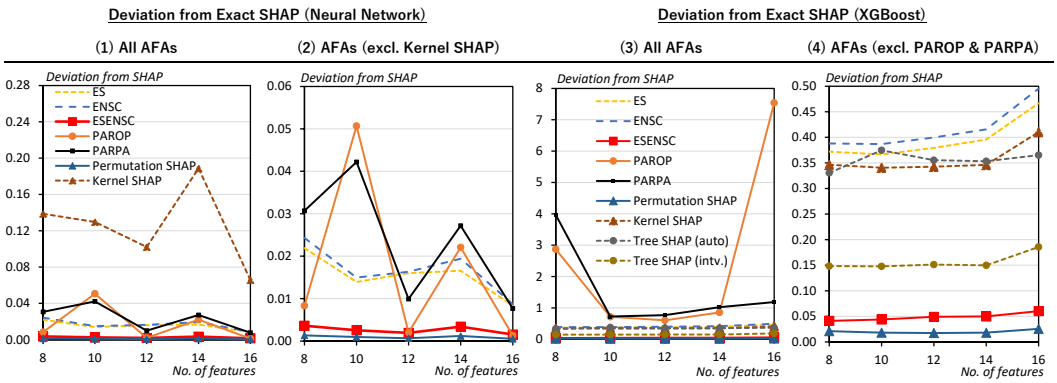


Figure 1: NAD values (deviation from exact SHAP) on the California Housing dataset. The left two panels correspond to Neural Network and the right two panels correspond to XGBoost.

4.1 Deviations from (Exact) SHAP

We first evaluate AFAs in terms of the deviation from exact SHAP using multiple complementary metrics that capture both magnitude differences and ranking consistency.

The first measure is Normalized Attribution Distance (NAD) from exact SHAP. For AFA ψ^{AFA},

$$\mathrm{NAD}(\psi^{AFA}) = \frac{\sum_{\tau=1}^{t}\sum_{j=1}^{n}\left|\psi_{\tau j}^{AFA} - \psi_{\tau j}^{SHAP}\right|}{\sum_{\tau=1}^{t}\left|v_\tau(N) - v_\tau(\emptyset)\right|}. \tag{3}$$

NAD captures the overall difference in attribution magnitudes, normalized by the scale of predictions.

The next measure is Mean Rank Correlation (MRC) with exact SHAP, defined as

$$\mathrm{MRC}(\psi^{AFA}) = \frac{1}{t}\sum_{\tau=1}^{t}\rho_\tau^{(\psi^{AFA},\psi^{SHAP})}, \tag{4}$$

where $\rho_\tau^{(\psi^{AFA},\psi^{SHAP})}$ denotes the Pearson correlation of the within-observation ranks of $\{\psi_{\tau j}^{AFA}\}_{j=1}^{n}$ and $\{\psi_{\tau j}^{SHAP}\}_{j=1}^{n}$. Together with NAD, this measure evaluates agreement with exact SHAP in both attribution magnitude and relative feature importance.

Figure 1 reports the NAD of AFAs where smaller values indicate closer agreement. The panels (1) and (2) correspond to the neural network (NN). In panel (1), which includes all AFAs, Kernel SHAP exhibits relatively large deviations from exact SHAP. The right panel, excluding Kernel SHAP, reveals clearer patterns: the PA–type AFAs (PAROP and PARPA) exhibit both large and volatile deviations, whereas the ES–type AFA ESENSC consistently achieves small deviations and strong approximation performance. The panels (3) and (4) show the results for XGBoost. The overall patterns are similar: PA–type AFAs again display large and unstable deviations, while ESENSC maintains consistently small deviations. ESENSC also performs favorably relative to TreeSHAP in this setting.

Why do PA–type AFAs exhibit such volatile behavior? Although PARPA was designed to address the order-reversal problem, it shows deviation patterns similar to PAROP, suggesting that additional sources of instability remain. A key mechanism arises when positive and negative contributions coexist. As an example, consider a two-player TU game with $v(\emptyset) = 0$, $v(1) = 10$, $v(2) = -9$, and $v(N) = 15$. Although no order reversal occurs, the proportional scaling factor $v(N)/(v(1) + v(2)) = 15$ becomes large because the denominator is close to zero, leading to extreme allocations ($\psi_1^{PA} = 150$, $\psi_2^{PA} = -135$). This illustrates that resolving order reversal alone does not prevent large distortions in PA–type methods.

Figure 2 reports the MRC of AFAs where higher values indicate closer agreement with exact SHAP. As in the case of NAD, panels (1) and (2) show the results for NN, while panels (3) and (4) for XGBoost. The insights are broadly consistent with those from the NAD results: PA–type AFAs exhibit large and volatile deviations from exact SHAP. In contrast, ESENSC closely approximates exact SHAP and achieves better performance than many SHAP approximation methods.

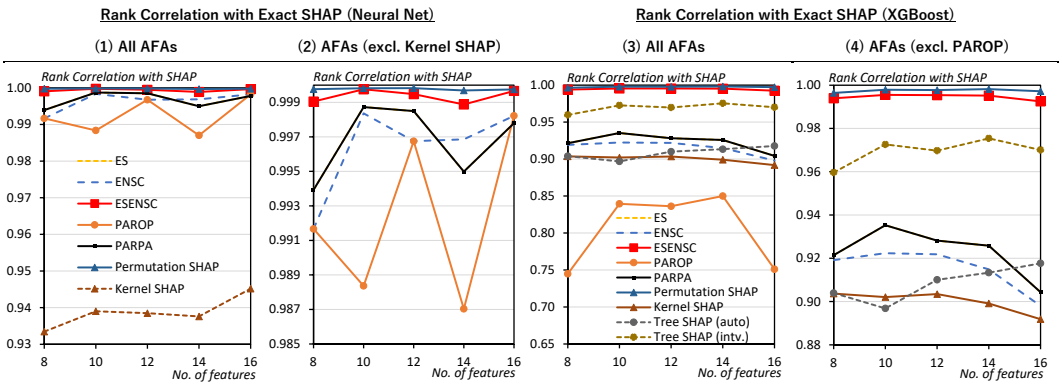


Figure 2: MRC values (rank correlation with exact SHAP) on the California Housing dataset. The left two panels correspond to Neural Network and the right two panels correspond to XGBoost.

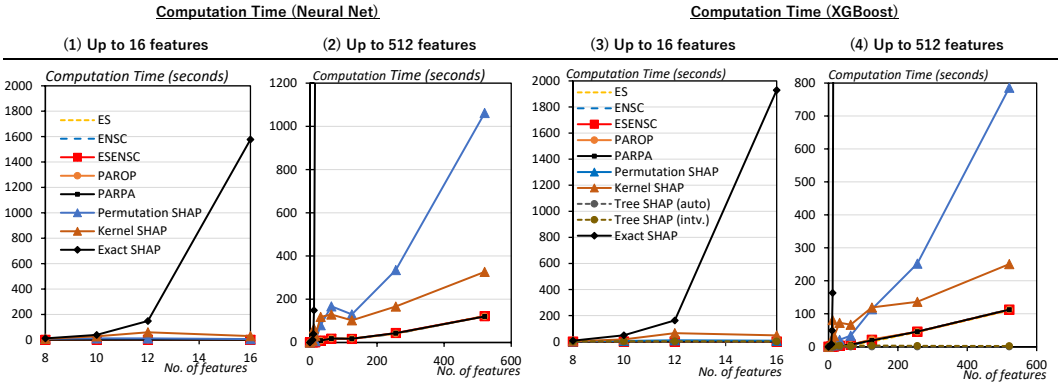


Figure 3: Computation time on the California Housing dataset. The left two panels correspond to Neural Network and the right two panels correspond to XGBoost.

4.2 Computation Time

We compare the computational cost of the attribution methods as a function of the number of features. Figure 3 shows the results for a neural network model, with panel (1) covering feature dimensions up to 16 where exact SHAP is computable. As expected, its computation time grows exponentially with the number of features, while all other AFA methods require substantially less time. Panel (2) extends the comparison to higher dimensions (up to 512 features), where computation of exact SHAP becomes infeasible. The results show that the proposed AFAs scale approximately linearly and remain computationally efficient even in high-dimensional settings. Their computation time is substantially lower than that of sampling-based methods such as Permutation SHAP and Kernel SHAP. The panels (3) and (4) report the results for XGBoost. The overall patterns are consistent with those for the neural network.

Although the runtime of sampling-based SHAP methods can be reduced by limiting the number of model evaluations, this typically worsens approximation accuracy, implying a trade-off between computational cost and precision. In contrast, the proposed AFAs are parameter-free once coalition values are specified, allowing direct and efficient computation.

Overall, under the evaluated settings, ESENSC offers a favorable trade-off between agreement with exact SHAP and computational cost, and compares favorably with the sampling-based SHAP approximations considered here.

5 Axiomatization of ESENSC

This section presents the theoretical foundations of ESENSC and its axiomatic characterization, showing that, for $n \neq 3$, it is uniquely determined by axioms tailored to XAI–TU games, analogous to classical Shapley characterizations.

ESENSC satisfies all five axioms in Theorem 1 for every n. The role of n can be summarized as follows. (i) For $n \leq 2$, every game is 2-additive, so ESENSC coincides with SHAP on $\mathbb{V}$ (Corollary 3 in Appendix B); the five axioms pin down this common rule, and the characterization does not distinguish ESENSC from SHAP. (ii) For $n = 3$, we have $2n + 2 = 2^n$, so information parsimony imposes no restriction; SHAP then also satisfies all five axioms, while ESENSC and SHAP differ whenever $\mathcal{A}(v) \neq N$ and $\lambda_N^v \neq 0$ (Appendix B). Hence uniqueness fails. (iii) For $n \geq 4$, information parsimony becomes binding ($2n+2 < 2^n$), and ESENSC is the unique rule satisfying the five axioms. Theorem 1 is stated for $n \neq 3$, but its substantive content lies in case (iii). For $n \geq 4$, the restriction to $\mathbb{V}$ is justified in Remark 1 and the related discussion in Appendix G.

To state the main result, we introduce the following definition.

Definition 1. *$S \subseteq N$ is relevant to ψ if there exist $v, w \in \mathbb{V}$ such that $v(T) = w(T)$ for any $T \subseteq N$ with $T \neq S$, and $\psi(v) \neq \psi(w)$.*

This definition captures the idea that a coalition S is relevant to a solution ψ if changing $v(S)$ alone, while keeping the worths of all other coalitions $T \subseteq N$ with $T \neq S$ unchanged, can alter the payoff vector assigned by ψ. In other words, S is relevant to ψ if its worth $v(S)$ can affect the payoffs determined by ψ. For example, every coalition $S \subseteq N$ is relevant to SHAP. In contrast, only $\emptyset$ and N are relevant to the equal division solution, which assigns $\frac{v(N)-v(\emptyset)}{n}$ to each player. Finally, no coalition is relevant to the zero solution, defined by $\psi_i(v) = 0$ for every $v \in \mathbb{V}$ and $i \in N$.

The main result of this section is the following:

Theorem 1. *Let $n \neq 3$. An attribution rule ψ satisfies the following five axioms—efficiency, the null player property, restricted differential marginality, pairwise interaction splitting, and information parsimony—if and only if $\psi = \psi^{\mathrm{ESENSC}}$.*

- ***Efficiency.*** *For any $v \in \mathbb{V}$, $\sum_{i \in N} \psi_i(v) = v(N) - v(\emptyset)$.*
- ***Null player property.*** *If $i \in N$ satisfies $v(S \cup i) = v(S)$ for any $S \subseteq N \setminus i$, then $\psi_i(v) = 0$.*
- ***Restricted differential marginality.*** *Let $v, w \in \mathbb{V}$ and $i, j \in N$ satisfy (i) $v(S \cup i) - v(S \cup j) = w(S \cup i) - w(S \cup j)$ for any $S \subseteq N \setminus \{i, j\}$, (ii) $i, j \in \mathcal{A}(v)$ and $i, j \in \mathcal{A}(w)$, and (iii) $\mathcal{A}(v) = \mathcal{A}(w)$. Then $\psi_i(v) - \psi_j(v) = \psi_i(w) - \psi_j(w)$.*
- ***Pairwise interaction splitting.*** *If $v \in \mathbb{V}$ is 2–additive, then $\psi_i(v) = \lambda_i^v + \sum_{j \neq i} \frac{\lambda_{\{i,j\}}^v}{2}$ for any $i \in N$, where λ^v denotes the Harsanyi dividends of v defined in Section 2.*
- ***Information parsimony.*** *$|\{S \subseteq N | S \text{ is relevant to } \psi\}| \leq 2n + 2$.*

As for the proof of Theorem 1, see Appendix G. Appendix H shows independence of the axioms in Theorem 1.

Here we compare the axioms in Theorem 1 for ESENSC with those used to axiomatize SHAP in [1], which are restated in Remark 2 in Appendix I. The first four axioms are all satisfied by SHAP. For $n \geq 4$, the distinction comes from the information parsimony axiom. Information parsimony limits the amount of coalition-value information on which the solution may depend. The particular bound $2n + 2$ can be interpreted as a simple linear information budget. When $n = 3$, we have $2n + 2 = 8 = 2^n$, so the axiom imposes no information restriction: all coalition values may be relevant. Taking the three-player case as the benchmark, the upper bound on the number of relevant coalition values increases by two for each additional player. This linear continuation gives $8 + 2(n - 3) = 2n + 2$. Thus, this axiom does not prescribe which coalition values should matter, nor does it require the actual number of relevant coalition values to increase with n; it only imposes a linear upper bound calibrated at the unrestricted three-player benchmark. This calibration is also sharp: Lemma 1 below shows that, for $n \geq 3$, the first four axioms require $2n + 2$ coalition values to be relevant.

The remaining four axioms are all compatible with SHAP. Efficiency and the null player property are minimal requirements for any attribution rule. For differential marginality, SHAP satisfies full differential marginality, requiring that identical marginal differences across coalitions imply identical attribution differences. In contrast, ESENSC satisfies only a restricted form of this axiom. More specifically, condition (ii) requires the two players being compared to be boundary-active in both games, while condition (iii) keeps the set of boundary-active players unchanged across the two games. Thus, restricted differential marginality compares payoff differences while holding fixed the set of players among whom the residual is redistributed.

Pairwise interaction splitting applies only to 2-additive games, in which all interactions are individual or pairwise. Within this restricted class, the axiom assigns each singleton dividend to the corresponding player and equally divides each pairwise dividend between the two players involved. The class of 2-additive games is the simplest extension of additive games in terms of interaction order that allows genuine interactions. By restricting the requirement to this class, the axiom imposes no direct restriction on the treatment of higher-order interactions and is therefore a relatively weak requirement.

Lemma 1. *Let ψ satisfy efficiency, the null player property, restricted differential marginality and pairwise interaction splitting. Then, every coalition in*

$$\mathcal{C} = \{S \subseteq N || S| = 0, 1, n-1, n\}$$

is relevant to ψ on $\mathbb{V}$.

Combining Lemma 1 with information parsimony yields the following boundary-information result.

Corollary 1. *If ψ satisfies the five axioms in Theorem 1, then for any $v, w \in \mathbb{V}$, if $v(S) = w(S)$ for $|S| = 0, 1, n-1$ and n, then $\psi(v) = \psi(w)$.*

We also use the following consequence of the five axioms in the proof of Theorem 1.

Corollary 2. *If ψ satisfies the five axioms in Theorem 1, then for any $v \in \mathbb{V}$ with $v(N) - v(\emptyset) = \sum_{j \in N} \frac{1}{2}\Big(v(j) - v(\emptyset) + v(N) - v(N \setminus j)\Big)$, it holds that $\psi_i(v) = \frac{1}{2}\Big(v(i) - v(\emptyset) + v(N) - v(N \setminus i)\Big)$ for any $i \in N$.*

6 Conclusion

This paper develops ESENSC, a computationally efficient and theoretically grounded alternative to SHAP for additive feature attribution. We demonstrate that ESENSC closely matches the attribution patterns of exact SHAP while achieving substantial computational gains, scaling linearly with the number of features. On the theoretical side, we establish an axiomatic characterization of ESENSC for $n \neq 3$, which is substantive for $n \geq 4$ (for $n \leq 2$, ESENSC coincides with SHAP), showing that it is uniquely determined by a set of properties tailored to XAI–TU games. This result shows how an explicit linear information constraint can yield substantial computational savings while retaining several axiomatic properties shared with SHAP. Empirically, under the evaluated settings, ESENSC performs competitively with, and in some settings better than, widely used SHAP approximation methods in terms of agreement with exact SHAP, while requiring substantially less computation and offering a theoretical foundation through its axiomatic characterization.

Overall, these results position ESENSC as a principled and practical alternative to SHAP, enabling reliable feature attribution for modern, high-dimensional models where exact computation is infeasible. An important direction for future research is to extend ESENSC beyond tabular data. Another important direction is to systematically understand the conditions under which ESENSC and SHAP yield different attribution patterns. This is particularly relevant because ESENSC focuses on marginal contributions at the boundary coalitions and may therefore underrepresent certain interaction effects that arise in intermediate coalitions.

Acknowledgments and Disclosure of Funding

We thank Yukihiko Funaki, Kiyoshi Izumi, Sungmin Kan, Shigeo Muto, Yuya Takada, Koji Takahashi, Makoto Yokoo and participants at European Meeting on Game Theory 2026 and Asian Game Theory Conference 2026 for their insightful comments and suggestions. The views expressed in this paper are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

This work was supported by JSPS KAKENHI Grant Numbers 25K05176 and 25K05018, and Nomura Foundation.

The authors declare no competing interests.

References

[1] A. Casajus. Differential marginality, van den Brink fairness, and the Shapley value. *Theory and Decision*, 71:163–174, 2011.

[2] C. Condevaux, S. Harispe, and S. Mussard. Fair and efficient alternatives to shapley-based attribution methods. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2022.

[3] I. Dragan, T. Driessen, and Y. Funaki. Collinearity between the Shapley value and the egalitarian division rules for cooperative games. *OR Spektrum*, 18:97–105, 1996.

[4] T. Driessen and Y. Funaki. Coincidence of and collinearity between game theoretic solutions. *OR Spektrum*, 13:15–30, 1991.

[5] J. C. Harsanyi. A simplified bargaining model for the n-person cooperative game. *International Economic Review*, 4(2):194–220, 1963.

[6] K. Hiraki, S. Ishihara, and J. Shino. Alternative methods to SHAP derived from properties of kernels: A note on theoretical analysis. In *Proceedings of the 2024 IEEE International Conference on Big Data (BigData 2024)*, pages 1321–1326. IEEE, 2024.

[7] T. Kongo. Balancing proportional viewpoints via ratio-based monotonicity. *Mimeo*, 2026.

[8] W. Li, G. Xu, and H. Sun. Maximizing the minimal satisfaction-characterizations of two proportional values. *Mathematics*, 8(1129), 2020.

[9] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1):56–67, 2020.

[10] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[11] H. Moulin. Implementing just and efficient decision-making. *Journal of Public Economics*, 16:193–213, 1981.

[12] H. Moulin. The separability axiom and equal-sharing methods. *Journal of Economic Theory*, 36:120–148, 1985.

[13] L. S. Shapley. A value for n-person games. *Annals of Mathematics Studies*, 28:307–318, 1953.

[14] P. D. Straffin and J. P. Heaney. Game theory and the Tennessee Valley Authority. *International Journal of Game Theory*, 10(1):35–43, 1981.

[15] E. Štrumbelj and I. Kononenko. An efficient explanation of individual classifications using game theory. *Journal of Machine Learning Research*, 11:1–18, 2010.

[16] E. Štrumbelj and I. Kononenko. Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, 41:647–665, 2014.

[17] R. van den Brink. An axiomatization of the Shapley value using a fairness property. *International Journal of Game Theory*, 30:309–319, 2002.

[18] R. van den Brink, Y. Chun, Y. Funaki, and Z. Zou. Balanced externalities and the proportional allocation of nonseparable contributions. *European Journal of Operational Research*, 307:975–983, 2023.

Appendix

A. Optimization interpretation of equal residual redistribution

This appendix formalizes the optimization interpretation discussed in Section 3.3.

Proposition 1 (Minimum-Euclidean efficiency correction)**.** *Suppose $v \in \mathbb{V}$. Among all attribution vectors $\psi \in \mathbb{R}^n$ that satisfy*

$$\sum_{i \in N} \psi_i = v(N) - v(\emptyset), \qquad \psi_i = 0 \quad \textit{for all } i \notin \mathcal{A}(v), \tag{5}$$

ESENSC defined in (1) is the unique solution to

$$\min_{\psi \in \mathbb{R}^n} \sum_{i \in \mathcal{A}(v)} \big(\psi_i - b_i(v)\big)^2 \quad \textit{subject to } (5). \tag{6}$$

Proof. Because $\psi_i = 0$ for $i \notin \mathcal{A}(v)$, the efficiency constraint is $\sum_{i \in \mathcal{A}(v)} \psi_i = v(N) - v(\emptyset)$. The Lagrangian for the boundary–active players is

$$\mathcal{L}(\psi, \mu) = \sum_{i \in \mathcal{A}(v)} \big(\psi_i - b_i(v)\big)^2 + \mu \left(\sum_{i \in \mathcal{A}(v)} \psi_i - v(N) + v(\emptyset) \right).$$

The first-order condition for each $i \in \mathcal{A}(v)$ is

$$2\big(\psi_i - b_i(v)\big) + \mu = 0,$$

so every boundary–active player receives the same adjustment, say c. Imposing efficiency gives

$$|\mathcal{A}(v)|c = v(N) - v(\emptyset) - \sum_{i \in \mathcal{A}(v)} b_i(v) = R(v),$$

which yields $c = R(v)/|\mathcal{A}(v)|$ and hence (1). The objective in (6) is strictly convex in the boundary–active players and the constraint set is affine, so this minimizer is unique. □

B. Harsanyi-dividend characterization of the ESENSC–SHAP difference

This appendix provides a complementary characterization of the difference between ESENSC and exact SHAP in terms of Harsanyi dividends λ_T^v defined in Section 2. Because $v(\emptyset)$ is absorbed into $\lambda_\emptyset^v$, no normalization of $v(\emptyset)$ is required. In particular, the analysis applies directly to XAI–TU games with nonzero baseline values. The Shapley attribution admits the standard dividend representation

$$\psi_i^{\mathrm{SHAP}}(v) = \sum_{T \subseteq N:\, i \in T} \frac{\lambda_T^v}{|T|}. \tag{7}$$

We refer to λ_T^v as the coalition-specific Harsanyi-dividend component associated with T and call $|T|$ its order. It captures the component of the game associated with coalition T after accounting for all proper subcoalitions. Components with $|T| \geq 3$ are referred to as higher-order Harsanyi-dividend components. Recall from Section 3.3 that players in $\mathcal{A}(v)$ are called boundary-active.

The results below are organized by the boundary-activity structure of v. Propositions 2 and 3 treat the case $\mathcal{A}(v) = N$, in which the residual is shared equally among all n players. Corollaries 3 and 4 hold for every $v \in \mathbb{V}$ without this assumption. Proposition 4 characterizes the boundary-inactive case $i \notin \mathcal{A}(v)$.

Proposition 2 (Exact ESENSC–SHAP difference when all players are boundary-active)**.** *Suppose that $v \in \mathbb{V}$ and $\mathcal{A}(v) = N$. Then, for every $i \in N$,*

$$\psi_i^{\mathrm{ESENSC}}(v) - \psi_i^{\mathrm{SHAP}}(v) = \sum_{\substack{T \subseteq N:\, i \in T \\ |T| \geq 2}} \left(\frac{1}{2} - \frac{1}{|T|}\right) \lambda_T^v + \frac{1}{n} \sum_{\substack{T \subseteq N \\ |T| \geq 2}} \left(1 - \frac{|T|}{2}\right) \lambda_T^v. \tag{8}$$

Equivalently, for $2 \leq \tau \leq n-1$, define

$$I(i,\tau) = \frac{1}{\binom{n-1}{\tau-1}} \sum_{\substack{T \subseteq N:\, i \in T \\ |T| = \tau}} \lambda_T^v, \qquad J(i,\tau) = \frac{1}{\binom{n-1}{\tau}} \sum_{\substack{T \subseteq N:\, i \notin T \\ |T| = \tau}} \lambda_T^v,$$

where $I(i,\tau)$ (resp. $J(i,\tau)$) is the average of the Harsanyi dividends of order τ that include (resp. do not include) i. Then

$$\psi_i^{\mathrm{ESENSC}}(v) - \psi_i^{\mathrm{SHAP}}(v) = \sum_{\tau=2}^{n-1} \binom{n-1}{\tau-1} \left(\frac{1}{\tau} - \frac{1}{n}\right) \left(1 - \frac{\tau}{2}\right) \big[J(i,\tau) - I(i,\tau)\big]. \tag{9}$$

In particular, the grand-coalition dividend cancels from the difference.

Proof. From the dividend representation in Section 2,

$$b_i(v) = \lambda_i^v + \frac{1}{2} \sum_{\substack{T \subseteq N:\, i \in T \\ |T| \geq 2}} \lambda_T^v. \tag{10}$$

Moreover,

$$\sum_{j \in N} b_j(v) = \sum_{j \in N} \lambda_j^v + \frac{1}{2} \sum_{\substack{T \subseteq N \\ |T| \geq 2}} |T| \lambda_T^v, \tag{11}$$

whereas

$$v(N) - v(\emptyset) = \sum_{\emptyset \neq T \subseteq N} \lambda_T^v. \tag{12}$$

Note that (10) to (12) hold for every $v \in \mathbb{V}$ and do not use the assumption $\mathcal{A}(v) = N$. Because $\mathcal{A}(v) = N$, the residual is redistributed over all n players. Therefore,

$$\psi_i^{\mathrm{ESENSC}}(v) = \lambda_i^v + \frac{1}{2} \sum_{\substack{T \subseteq N:\, i \in T \\ |T| \geq 2}} \lambda_T^v + \frac{1}{n} \sum_{\substack{T \subseteq N \\ |T| \geq 2}} \left(1 - \frac{|T|}{2}\right) \lambda_T^v.$$

Subtracting (7) gives (8). Regrouping the terms in (8) by coalition size and using $\binom{n-1}{\tau-1}/\binom{n-1}{\tau} = \tau/(n-\tau)$ yields (9). For $T = N$, the coefficient is $(1/2 - 1/n) + (1/n)(1 - n/2) = 0$, so the grand-coalition dividend cancels. □

Equation (9) makes the role of Harsanyi-dividend order explicit. The factor $1 - \tau/2$ is zero at $\tau = 2$, so second-order Harsanyi-dividend components do not generate an ESENSC–SHAP difference. More generally, if $I(i,\tau) = J(i,\tau)$ at every relevant higher order, then the two attributions coincide for feature i even when higher-order Harsanyi-dividend components are present.

Proposition 3 (Bound on deviations induced by higher-order dividend components)**.** *Suppose again that $\mathcal{A}(v) = N$ and define $\Delta_i = \psi_i^{\mathrm{ESENSC}}(v) - \psi_i^{\mathrm{SHAP}}(v)$. Let*

$$\mathcal{H}_i^+ = \{T \subseteq N : i \in T,\ 3 \leq |T| \leq n-1\}, \qquad \mathcal{H}_i^- = \{T \subseteq N : i \notin T,\ 3 \leq |T| \leq n-1\}.$$

Then

$$\Delta_i = \sum_{T \in \mathcal{H}_i^+} \frac{(|T|-2)(n-|T|)}{2|T|n} \lambda_T^v - \sum_{T \in \mathcal{H}_i^-} \frac{|T|-2}{2n} \lambda_T^v. \tag{13}$$

Hence,

$$|\Delta_i| \le \sum_{T \in \mathcal{H}_i^+} \frac{(|T|-2)(n-|T|)}{2|T|n} |\lambda_T^v| + \sum_{T \in \mathcal{H}_i^-} \frac{|T|-2}{2n} |\lambda_T^v|. \tag{14}$$

If $\mathcal{H} = \{T \subseteq N : 3 \le |T| \le n-1\}$*, then the simpler bound*

$$\left|\psi_i^{\mathrm{ESENSC}}(v) - \psi_i^{\mathrm{SHAP}}(v)\right| \le \frac{1}{2} \sum_{T \in \mathcal{H}} |\lambda_T^v| \tag{15}$$

holds for every $i \in N$*.*

Proof. Equation (13) follows by collecting the coefficients in (8). The terms of order two vanish and the grand-coalition term cancels, leaving only coalitions of sizes $3, \dots, n-1$. The first inequality follows from the triangle inequality. Both coefficients in (14) are at most $1/2$, which yields (15). □

The bound is necessarily scale-dependent: multiplying the entire game by a constant proportionally scales the Harsanyi dividends and both attribution vectors. Its implication is therefore relative rather than absolute—ESENSC must remain close to SHAP whenever the total absolute mass of the higher-order Harsanyi-dividend components associated with intermediate coalitions is small.

The assumption $\mathcal{A}(v) = N$ enters only through the denominator of the residual share. For general $v \in \mathbb{V}$ and $i \in \mathcal{A}(v)$, the same argument yields (8) with $1/n$ replaced by $1/|\mathcal{A}(v)|$. In that case the coefficient on λ_N^v equals $(n-2)(|\mathcal{A}(v)|-n)/(2n|\mathcal{A}(v)|)$, which is nonzero whenever $|\mathcal{A}(v)| < n$. The grand-coalition dividend then no longer cancels, and the bound (15) need not hold. If every player outside $\mathcal{A}(v)$ is a null player, Propositions 2 and 3 continue to hold with N and n replaced by $\mathcal{A}(v)$ and $|\mathcal{A}(v)|$.

Corollary 3 (Exact agreement on 2-additive games)**.** *For every 2-additive game* $v \in \mathbb{V}$*, ESENSC and exact SHAP coincide:* $\psi^{\mathrm{ESENSC}}(v) = \psi^{\mathrm{SHAP}}(v)$*.*

Proof. In a 2-additive game, $\lambda_T^v = 0$ for all $|T| \ge 3$. Hence

$$b_i(v) = \lambda_i^v + \frac{1}{2} \sum_{\substack{T \ni i \\ |T|=2}} \lambda_T^v = \psi_i^{\mathrm{SHAP}}(v).$$

Furthermore, $\sum_{i \in N} b_i(v) = v(N) - v(\emptyset)$, so the ESENSC residual is zero. Thus the equality holds for both boundary-active and boundary-inactive players. □

Corollary 4 (Positivity on monotone 2-additive games)**.** *If* $v \in \mathbb{V}$ *is monotone (i.e.,* $v(S) \le v(T)$ *whenever* $S \subseteq T$*) and 2-additive, then* $\psi_i^{\mathrm{ESENSC}}(v) \ge 0$ *for every* $i \in N$*.*

Proof. By Corollary 3, ESENSC coincides with the Shapley value. For a monotone game, every marginal contribution entering the Shapley value is nonnegative. Hence all ESENSC attributions are nonnegative. □

Positivity does not hold in general, even for monotone games. For example, let $N = \{1,2,3,4\}$, let $v(\emptyset) = 0$, let $v(N) = v(\{2,3,4\}) = 1$, let $v(S) = 0.1$ for every coalition containing player 1 except N, and let $v(S) = 0$ otherwise. This game is monotone, yet ESENSC assigns player 1 the negative attribution -0.05.

Proposition 4 (Characterization of false-null attribution)**.** *Suppose that* $v \in \mathbb{V}$ *and* $i \notin \mathcal{A}(v)$*, so that ESENSC assigns* $\psi_i^{\mathrm{ESENSC}}(v) = 0$*. Then*

$$\lambda_i^v = 0, \qquad \sum_{T \subseteq N:\, i \in T} \lambda_T^v = 0. \tag{16}$$

Exact SHAP nevertheless assigns a nonzero value to i *if and only if*

$$\sum_{T \subseteq N:\, i \in T} \frac{\lambda_T^v}{|T|} \neq 0. \tag{17}$$

Consequently, a false-null judgment relative to exact SHAP can occur only when nonzero dividends involving i appear at different coalition cardinalities and cancel in the unweighted boundary sum in (16)*, but not under the Shapley weights in* (17)*.*

Proof. The condition $i \notin \mathcal{A}(v)$ means $v(i) = v(\emptyset)$ and $v(N) = v(N \setminus i)$. The first equality gives $\lambda_i^v = 0$, while the second gives $\sum_{T \ni i} \lambda_T^v = 0$. Equation (7) then yields the stated condition for a nonzero SHAP value. If all nonzero dividends involving i had the same cardinality, the unweighted and Shapley-weighted sums would differ only by a common positive factor, so the two cancellation conditions would coincide. □

In particular, when $n = 3$, ESENSC and SHAP coincide if and only if $\mathcal{A}(v) = N$ or $\lambda_N^v = 0$; hence, they differ whenever $\mathcal{A}(v) \neq N$ and $\lambda_N^v \neq 0$.

C. Derivation of PARPA

In this appendix, we derive PARPA and discuss why PARPA avoids the order-reversal problem. First, we introduce a complementary proportional rule, RPA, which reverses the weighting structure of PA. While PA assigns weights $v(j) - v(\emptyset)$, RPA assigns $v(N) - v(\emptyset) - \sum_{i \neq j}(v(i) - v(\emptyset))$, the residual value that remains when the marginal contributions of all other features are deducted from the total surplus. Formally,

$$\begin{aligned} \psi_j^{\mathrm{RPA}} &= \frac{v(N) - v(\emptyset) - \sum_{i \neq j} \{v(i) - v(\emptyset)\}}{\sum_{i \in N} \left\{ v(N) - v(\emptyset) - \sum_{k \neq i} \{v(k) - v(\emptyset)\} \right\}} \cdot \{v(N) - v(\emptyset)\} \\ &= \frac{v(N) - v(\emptyset) - \sum_{i \neq j} \{v(i) - v(\emptyset)\}}{n\{v(N) - v(\emptyset)\} - (n-1) \sum_{i \in N} \{v(i) - v(\emptyset)\}} \cdot \{v(N) - v(\emptyset)\}. \end{aligned} \tag{18}$$

In Example 1 in Subsection 3.4, we have $\psi_1^{\mathrm{RPA}} = -1/3$ and $\psi_2^{\mathrm{RPA}} = 4/3$. Thus, RPA satisfies the order preservation property in this case. This suggests that when the signs of $v(N) - v(\emptyset)$ and $\sum_{i \in N}\{v(i) - v(\emptyset)\}$ differ, RPA may be preferable to PA.

We now generalize this argument. Assume $v(N) - v(\emptyset) \neq 0$, since otherwise attribution is trivial. From (2) and (18), the order-reversal problem arises in PA when $(v(N) - v(\emptyset))/\left[\sum_{i \in N} \{v(i) - v(\emptyset)\}\right] < 0$, and arises in RPA when $(v(N) - v(\emptyset))/\left[n\{v(N) - v(\emptyset)\} - (n-1)\sum_{i \in N} \{v(i) - v(\emptyset)\}\right] < 0$. In both cases, whether order reversal occurs depends solely on the sign combination of $v(N) - v(\emptyset)$ and $\sum_{i \in N}\{v(i) - v(\emptyset)\}$.

Table 1 summarizes the cases, where the first (second) entry in each cell corresponds to PA (RPA). "OK" indicates no order reversal, "X" indicates it necessarily arises, "?" indicates it may arise, and "NA" indicates the rule is undefined. The table shows that for any sign configuration, at least one of PA or RPA avoids order reversal. Thus, by selecting PA or RPA appropriately, the order-reversal problem can be avoided.

$v(N) - v(\emptyset)$ \ $\sum_{i \in N}\{v(i) - v(\emptyset)\}$	Positive	Negative	Zero
Positive	(OK, ?)	(X, OK)	(NA, OK)
Negative	(X, OK)	(OK, ?)	(NA, OK)

Table 1: PA and RPA: cases where the order-reversal problem does not arise ("OK")

A simple selection rule is to choose the rule that corresponds to the "OK" case:

$$\psi_j^{\mathrm{PARPA}} = \begin{cases} \psi_j^{\mathrm{PA}} & \text{if } \left[v(N) - v(\emptyset)\right] \cdot \left[\sum_{i \in N}\{v(i) - v(\emptyset)\}\right] > 0, \\ \psi_j^{\mathrm{RPA}} & \text{otherwise.} \end{cases}$$

In sum, the rule ψ^{PARPA} is a PA–type AFA to address the order-reversal problem and preserves the ranking induced by $v(j) - v(\emptyset)$.

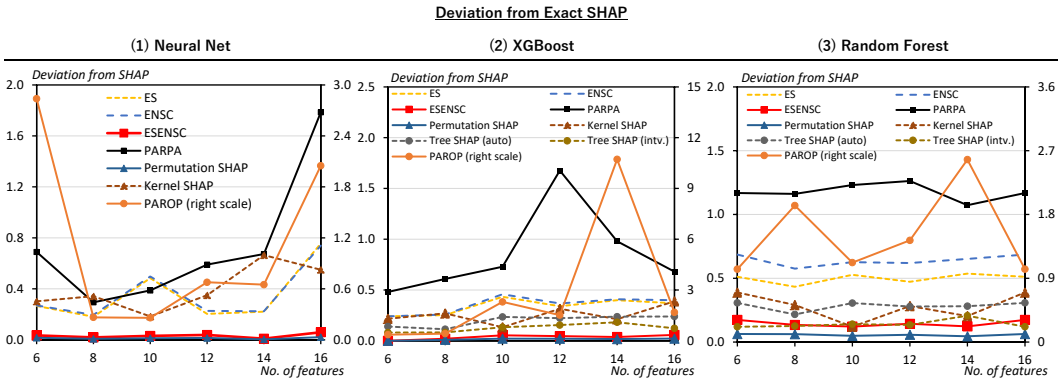


Figure 4: NAD values (deviation from exact SHAP) on the HELOC dataset. The three panels correspond to Neural Network (left), XGBoost (center), and Random Forest (right).

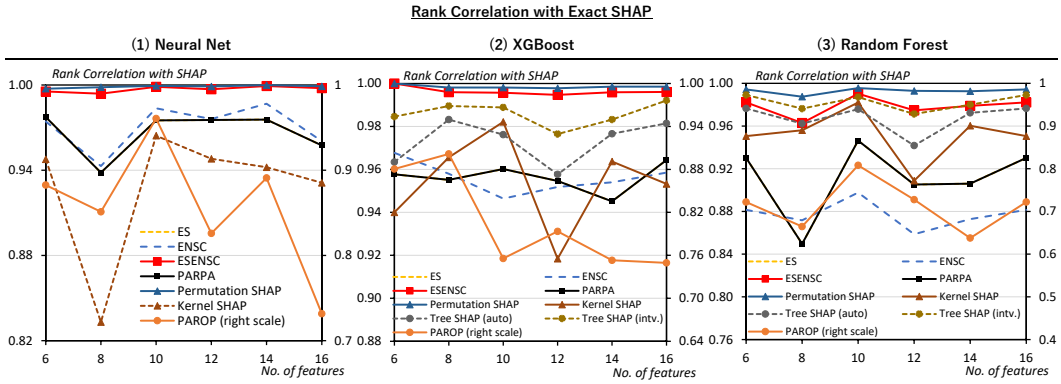


Figure 5: MRC values (rank correlation with exact SHAP) on the HELOC dataset. The three panels correspond to Neural Network (left), XGBoost (center), and Random Forest (right).

D. Additional Experiments on the HELOC Dataset

This appendix presents additional experiments using the HELOC dataset to further validate the findings in the main text. In this experiment, formulated as a regression task, the target variable is *ExternalRiskEstimate*, a consolidated risk score derived from credit bureau data. The dataset contains 22 features in total; to evaluate the performance of AFA methods across various input dimensions, we select n features, where $n \in \{6, 8, 10, 12, 14, 16\}$. In addition to the Neural Network and XGBoost models employed in the main text, we include Random Forest as an additional predictive model to ensure a broader evaluation. For the random forest model, we use `RandomForestRegressor` from scikit-learn and tune its hyperparameters via grid search with five-fold cross-validation. The search space includes the number of trees (50, 100, and 200) and the maximum tree depth (10, 20, or unrestricted). We select the model that minimizes cross-validated mean squared error. Consistent with the main experiments, we use the first 1,000 observations and compute attributions for the test split (see Appendix E). Each AFA method is assessed based on its deviation from the exact SHAP values, using the Normalized Attribution Distance (NAD) and the Mean Rank Correlation (MRC) defined in Eqs. (3) and (4), respectively.

The experimental results are summarized in Figures 4 and 5, which correspond to the NAD and MRC metrics. In each figure, the left, center, and right panels show the results for the Neural Network, XGBoost, and Random Forest models, respectively. The horizontal axis of each panel represents the number of features, while the vertical axis indicates the value of each deviation indicator. Note that the values for PAROP are displayed on the right-hand scale, as their magnitude differs significantly from those of other AFA methods.

These results strongly support the robustness of the findings presented in Section 4. Specifically, ESENSC consistently exhibits a minimal deviation from the exact SHAP values, with its rank

correlation remaining remarkably close to 1. Furthermore, ESENSC generally outperforms existing SHAP approximation algorithms in terms of overall accuracy. In contrast, PA–type AFA methods show unstable performance; in particular, PAROP—which lacks the order preservation property—demonstrates significant instability. These patterns are similar across the datasets, predictive models, and feature dimensions considered, and are consistent with the theoretical analysis provided in the main text.

E. Model Description

In Section 4, we adopt XGBoost and a neural network as the learning models. In this appendix, we describe these models in detail.

For XGBoost, we use the `XGBRegressor` implementation and tune its hyperparameters via randomized search with three-fold cross-validation. The search space includes the number of estimators (100–1000), learning rate (uniformly sampled from 0 to 0.5), maximum tree depth (2 to 10), subsample ratio (0.5 to 1.0), and column sampling ratio (0.8–1.0). We perform 200 iterations and select the model that minimizes cross-validated mean squared error. These ranges follow common practice in empirical applications and provide a flexible yet standard tuning setup.

For the neural network model, we use `MLPRegressor` from scikit-learn within a preprocessing pipeline that includes feature standardization. Hyperparameters are tuned via grid search with five-fold cross-validation. We consider two hidden-layer architectures with moderate widths (ranging from $(64, 64)$ to $(256, 128)$), ReLU and tanh activation functions, L2 regularization parameters (α) between 10^{-7} and 10^{-4}, and initial learning rates of 10^{-3} and 3×10^{-4}. Training is performed with a maximum of 5000 iterations and early stopping based on a validation split. These choices follow common practice for multilayer perceptrons in tabular regression. For each dataset, we use the first 1,000 observations in the original order of the data. These are split into training (80%) and test (20%) sets, and attributions are computed for the test observations. A fixed random seed (42) is used for data generation, the train/test split, and model training.

Overall, both models are trained with cross-validated hyperparameter tuning over standard parameter ranges, ensuring that the comparison of attribution methods is conducted under conventional and well-established settings.

F. Proofs of Lemma 1 and Corollaries 1 and 2

Proof of Lemma 1: The proof consists of three claims.

Claim 1. *Each of $\emptyset$ and N is relevant to ψ on $\mathbb{V}$.*

Proof of Claim 1: Let $v : 2^N \to \mathbb{R}$ be $v(\emptyset) = 0$ and $v(S) = 1$ if $S \neq \emptyset$. Let $w : 2^N \to \mathbb{R}$ be $w(\emptyset) = -1$ and $w(S) = v(S)$ if $S \neq \emptyset$. Let $u : 2^N \to \mathbb{R}$ be $u(N) = 2$ and $u(S) = v(S)$ if $S \neq N$. Now $v, w, u \in \mathbb{V}$.

From efficiency, $\sum_{i \in N} \psi_i(v) = 1$ and $\sum_{i \in N} \psi_i(w) = 2$. Since v and w differ only in the worth of $\emptyset$, $\emptyset$ is relevant to ψ on $\mathbb{V}$. Similarly, $\sum_{i \in N} \psi_i(u) = 2$, and since v and u differ only in the worth of N, N is relevant to ψ on $\mathbb{V}$. □

If $n = 1$, Claim 1 is sufficient. If $n = 2$, every game is 2-additive. Hence, pairwise interaction splitting implies that $\psi_i(v) = \frac{1}{2}(v(i) - v(\emptyset) + v(N) - v(N \setminus i))$ for any $v \in \mathbb{V}$ and $i \in N$, which implies the desired result. In what follows, we consider the case when $n \geq 3$.

Claim 2. *Fix $i \in N$. Then, $\{i\}$ is relevant to ψ on $\mathbb{V}$.*

Proof of Claim 2: Let $v : 2^N \to \mathbb{R}$ be $v(\emptyset) = 0$ and $v(S) = 1$ if $S \neq \emptyset$. Let $w : 2^N \to \mathbb{R}$ be $w(S) = |S|$ for any $S \subseteq N$. Let $u : 2^N \to \mathbb{R}$ be $u(\emptyset) = u(i) = 0$ and $u(S) = 1$ if $S \neq \emptyset, \{i\}$. Now $v, w, u \in \mathbb{V}$.

Because w is 2-additive, $\psi_k(w) = 1$ for any $k \in N$. It is straightforward to verify that v, w and any $j, k \in N$ satisfy conditions (i)-(iii) of restricted differential marginality. From restricted differential marginality, $\psi_k(v) - \psi_j(v) = \psi_k(w) - \psi_j(w) = 0$ for any $k, j \in N$. From efficiency, $\psi_k(v) = \frac{v(N) - v(\emptyset)}{n} = \frac{1}{n}$ for any $k \in N$. Because i is a null player in u, the null player property implies that $\psi_i(u) = 0$. Since v and u differ only in the worth of $\{i\}$, while $\psi_i(v) = \frac{1}{n} \neq 0 = \psi_i(u)$, $\{i\}$ is relevant to ψ on $\mathbb{V}$. □

Claim 3. *Fix $i \in N$. Then, $N \setminus i$ is relevant to ψ on $\mathbb{V}$.*

Proof of Claim 3: Let $v : 2^N \to \mathbb{R}$ be $v(N) = 1$ and $v(S) = 0$ if $S \neq N$. Let $w : 2^N \to \mathbb{R}$ be $w(S) = |S|$ for any $S \subseteq N$. Let $u : 2^N \to \mathbb{R}$ be $u(N) = u(N \setminus i) = 1$ and $u(S) = 0$ if $S \neq N, N \setminus \{i\}$. Now $v, w, u \in \mathbb{V}$.

Because w is 2-additive, $\psi_k(w) = 1$ for any $k \in N$. It is straightforward to verify that v, w and any $j, k \in N$ satisfy conditions (i)-(iii) of restricted differential marginality. From restricted differential marginality, $\psi_k(v) - \psi_j(v) = \psi_k(w) - \psi_j(w) = 0$ for any $k, j \in N$. From efficiency, $\psi_k(v) = \frac{v(N)-v(\emptyset)}{n} = \frac{1}{n}$ for any $k \in N$. Because i is a null player in u, the null player property implies that $\psi_i(u) = 0$. Since v and u differ only in the worth of $N \setminus \{i\}$, while $\psi_i(v) = \frac{1}{n} \neq 0 = \psi_i(u)$, $N \setminus \{i\}$ is relevant to ψ on $\mathbb{V}$. □

Note that Claims 2 and 3 hold for any $i \in N$. From Claims 1 to 3, Lemma 1 is obtained. □

Proof of Corollary 1: If $n \geq 3$, $|\mathcal{C}| = 2n + 2$. For $n = 1, 2, 3$, $\mathcal{C} = 2^N$, and hence the following conclusion is immediate. Suppose $n \geq 4$. Since information parsimony allows at most $2n + 2$ relevant coalitions, Lemma 1 implies that no coalition outside $\mathcal{C}$ is relevant to ψ. Let $v, w \in \mathbb{V}$ satisfy $v(S) = w(S)$ for every $S \in \mathcal{C}$, and enumerate the coalitions $2^N \setminus \mathcal{C}$ as $S_1, \ldots, S_m$. Starting from v, construct a sequence of games $v^0, \ldots v^m$, where $v^0 = v$ and v^k is obtained from v^{k-1} by replacing the worth of S_k with $w(S_k)$. Since membership in $\mathbb{V}$ depends only on the worths of coalitions in $\mathcal{C}$, every v^k belongs to $\mathbb{V}$. Moreover, since S_k is not relevant to ψ, we have $\psi(v^{k-1}) = \psi(v^k)$ for every $k = 1, \ldots m$. Therefore, $\psi(v) = \psi(v^0) = \cdots = \psi(v^m) = \psi(w)$. □

Proof of Corollary 2: Let $v \in \mathbb{V}$ satisfy the condition in Corollary 2. If $n \leq 2$, every game is 2-additive, and the result follows directly from pairwise interaction splitting. Suppose $n \geq 3$. For each $i \in N$, let

$$d_i = v(N) - v(N \setminus i) - v(i) + v(\emptyset).$$

Construct a 2-additive game w whose Harsanyi dividends are given by

$$\lambda^w_\emptyset = v(\emptyset), \qquad \lambda^w_i = v(i) - v(\emptyset),$$

and, for distinct $i, j \in N$,

$$\lambda^w_{\{i,j\}} = \frac{d_i + d_j - \frac{1}{n-1}\sum_{k\in N} d_k}{n-2}$$

with all remaining Harsanyi dividends equal to zero. Then

$$\sum_{j\neq i} \lambda^w_{\{i,j\}} = d_i$$

for every $i \in N$. Hence,

$$w(i) = v(i) \quad \text{and} \quad w(N) - w(N \setminus i) = v(N) - v(N \setminus i).$$

Moreover,

$$w(N) - w(\emptyset) = \sum_{i\in N} \lambda^w_i + \frac{1}{2}\sum_{i\in N} d_i = \sum_{i\in N} \frac{1}{2}\big(v(i) - v(\emptyset) + v(N) - v(N \setminus i)\big) = v(N) - v(\emptyset).$$

Thus, $w(N) = v(N)$ and $w(N \setminus i) = v(N \setminus i)$, which imply $w \in \mathbb{V}$. By Corollary 1, $\psi(v) = \psi(w)$. Since w is 2-additive, pairwise interaction splitting implies

$$\psi_i(v) = \psi_i(w) = \lambda^w_i + \frac{1}{2}\sum_{j\neq i} \lambda^w_{\{i,j\}} = \frac{1}{2}\big(v(i) - v(\emptyset) + v(N) - v(N \setminus i)\big)$$

for every $i \in N$. □

G. Proof of Theorem 1

This appendix shows the proof of Theorem 1. First, we prove the "if" part with the following Claim 4:

Claim 4. *ψ^{ESENSC} satisfies efficiency, the null player property, restricted differential marginality, pairwise interaction splitting, and information parsimony.*

Proof of Claim 4: Efficiency and the null player property are apparent. Regarding restricted differential marginality, let $v, w \in \mathbb{V}$ and $i, j \in N$ satisfying (i),(ii), and (iii) in the definition of the axiom. Then,

$$\psi_i^{\mathrm{ESENSC}}(v) - \psi_j^{\mathrm{ESENSC}}(v) = b_i(v) - b_j(v) = \frac{1}{2}(v(i) - v(j) - v(N \setminus i) + v(N \setminus j))$$
$$= \frac{1}{2}(w(i) - w(j) - w(N \setminus i) + w(N \setminus j)) = b_i(w) - b_j(w) = \psi_i^{\mathrm{ESENSC}}(w) - \psi_j^{\mathrm{ESENSC}}(w).$$

Regarding information parsimony, ψ^{ESENSC} depends only on coalition values in $\mathcal{C}$ and $|\mathcal{C}| \leq 2n+2$ for every n. Pairwise interaction splitting follows directly from Corollary 3 and its proof. □

We then prove the "Only-if" part with the following Claim 5 and Claim 6. Let ψ be an attribution rule satisfying the five axioms in Theorem 1.

Claim 5. *From the null player property and Corollary 1, for any $v \in \mathbb{V}$ and $i \in N$ with $v(i) = v(\emptyset)$ and $v(N) = v(N \setminus i)$, $\psi_i(v) = 0$.*

Proof of Claim 5: If $n = 1$ or 2, the null player property implies the fact. If $n \geq 4$, given $v \in \mathbb{V}$ and $i \in N$ with $v(i) = v(\emptyset)$ and $v(N) = v(N \setminus i)$, let $v^i : 2^N \to \mathbb{R}$, where

$$v^i(S) = \begin{cases} v(S) & \text{if } S \ni i \text{ and } |S| \neq 2 \\ v(S \setminus i) & \text{if } S \ni i \text{ and } |S| = 2 \\ v(S) & \text{if } S \not\ni i \text{ and } |S| = 0, 1, n-1 \\ v(S \cup i) & \text{if } S \not\ni i \text{ and } |S| \neq 0, 1, n-1. \end{cases}$$

By definition, $v^i(\emptyset) = v(\emptyset)$, for any $j \in N$, $v^i(j) = v(j)$ and $v^i(N \setminus j) = v(N \setminus j)$,[10] and $v^i(N) = v(N)$. Hence, $v^i \in \mathbb{V}$. From Corollary 1,

$$\psi(v^i) = \psi(v). \tag{19}$$

Additionally, by definition, for any $S \subseteq N \setminus i$,

$$v^i(S \cup i) - v^i(S) = \begin{cases} v(i) - v(\emptyset) = 0 & \text{if } |S| = 0, \\ v(S) - v(S) = 0 & \text{if } |S| = 1, \\ v(S \cup i) - v(S \cup i) = 0 & \text{if } |S| = 2, \ldots, n-2, \\ v(N) - v(N \setminus i) = 0 & \text{if } |S| = n-1. \end{cases}$$

Thus, i is a null player in v^i. From the null player property,

$$\psi_i(v^i) = 0. \tag{20}$$

Combining (19) and (20), we obtain the desired result. □

Claim 6. *If ψ satisfies the five axioms in Theorem 1, then $\psi = \psi^{\mathrm{ESENSC}}$.*

Proof of Claim 6: We divide the proof into two cases.

Case (i): $v \in \mathbb{V}$ with $v(N) - v(\emptyset) = \sum_{k \in N} \frac{1}{2}(v(k) - v(\emptyset) + v(N) - v(N \setminus k))$. Equivalently, Case (i) is characterized by $R(v) = 0$. Note that if $n = 1$ or 2, any $v \in \mathbb{V}$ belongs to this case. From Corollary 2, the desired result is obtained.

Case (ii): $v \in \mathbb{V}$ with $v(N) - v(\emptyset) \neq \sum_{k \in N} \frac{1}{2}(v(k) - v(\emptyset) + v(N) - v(N \setminus k))$. Note that this case occurs only when $n \geq 3$. Under the maintained assumption $n \neq 3$, we therefore have $n \geq 4$ throughout Case (ii). Note also that from the definition of $\mathbb{V}$, $\mathcal{A}(v) \neq \emptyset$.

We further subdivide this case into two cases.

Case (ii)-(a): $|\mathcal{A}(v)| = 1$. From Claim 5, $\psi_\ell(v) = 0 = \psi_\ell^{\mathrm{ESENSC}}(v)$ for any $\ell \in N \setminus \mathcal{A}(v)$. From efficiency, for $i \in \mathcal{A}(v)$,

$$\psi_i(v) = v(N) - v(\emptyset) = b_i(v) + \frac{v(N) - v(\emptyset) - b_i(v)}{1} = b_i(v) + \frac{R(v)}{|\mathcal{A}(v)|} = \psi_i^{\mathrm{ESENSC}}(v).$$

[10] Note that because $n \geq 4$, it holds that $n - 1 \neq 2$.

Case (ii)-(b): $|\mathcal{A}(v)| \geq 2$. Given $v \in \mathbb{V}$, let $\bar{v} : 2^N \to \mathbb{R}$ where

$$\bar{v}(S) = \begin{cases} v(S) & |S| = 0, \ldots, n-2, \\ v(S) - R(v) & |S| = n-1, n. \end{cases}$$

Because $R(v) \neq 0$, $v \neq \bar{v}$. Additionally, because $n \geq 3$, for any $\ell \in N$, $\bar{v}(\ell) - \bar{v}(\emptyset) = v(\ell) - v(\emptyset)$ and $\bar{v}(N) - \bar{v}(N \setminus \ell) = v(N) - v(N \setminus \ell)$. Hence, $\bar{v} \in \mathbb{V}$, $b_\ell(\bar{v}) = b_\ell(v)$ for any $\ell \in N$, and $\mathcal{A}(\bar{v}) = \mathcal{A}(v)$. Moreover, for any $i, j \in \mathcal{A}(v)$ and any $S \subseteq N \setminus \{i, j\}$, $\bar{v}(S \cup i) - \bar{v}(S \cup j) = v(S \cup i) - v(S \cup j)$, since $S \cup i$ and $S \cup j$ have the same cardinality. Additionally,

$$R(\bar{v}) = \bar{v}(N) - \bar{v}(\emptyset) - \sum_{i \in \mathcal{A}(\bar{v})} b_i(\bar{v}) = v(N) - R(v) - v(\emptyset) - \sum_{i \in \mathcal{A}(v)} b_i(v) = 0,$$

Hence, $\bar{v}$ belongs to Case (i). From Case (i),

$$\psi(\bar{v}) = \psi^{\mathrm{ESENSC}}(\bar{v}).$$

From restricted differential marginality, for any $i, j \in \mathcal{A}(v)$,

$$\begin{aligned} \psi_i(v) - \psi_j(v) = \psi_i(\bar{v}) - \psi_j(\bar{v}) = \psi_i^{\mathrm{ESENSC}}(\bar{v}) - \psi_j^{\mathrm{ESENSC}}(\bar{v}) = b_i(\bar{v}) - b_j(\bar{v}) \\ = \frac{1}{2}\left(\bar{v}(i) - \bar{v}(j) - \bar{v}(N \setminus i) + \bar{v}(N \setminus j)\right) = \frac{1}{2}\left(v(i) - v(j) - v(N \setminus i) + v(N \setminus j)\right) \\ = b_i(v) - b_j(v), \end{aligned}$$

which implies that

$$\psi_i(v) - b_i(v) = \psi_j(v) - b_j(v). \tag{21}$$

Hence, every $i \in \mathcal{A}(v)$ receives the same adjustment, say c. From Claim 5,

$$\psi_\ell(v) = 0 = \psi_\ell^{\mathrm{ESENSC}}(v) \text{ for any } \ell \in N \setminus \mathcal{A}(v). \tag{22}$$

From efficiency and equation (22),

$$\sum_{j \in \mathcal{A}(v)} \psi_j(v) = v(N) - v(\emptyset). \tag{23}$$

From equations (21) and (23),

$$|\mathcal{A}(v)|c = v(N) - v(\emptyset) - \sum_{i \in \mathcal{A}(v)} b_i(v) = R(v),$$

and thus,

$$c = \frac{R(v)}{|\mathcal{A}(v)|}.$$

Therefore, for any $i \in \mathcal{A}(v)$,

$$\psi_i(v) = b_i(v) + c = \psi_i^{\mathrm{ESENSC}}(v).$$

Together with equation (22), $\psi(v) = \psi^{\mathrm{ESENSC}}(v)$. □
(The end of the proof of Theorem 1.)

The restriction of the axiomatization domain to $\mathbb{V}$ in Theorem 1, as compared with Remark 2 in the subsequent Appendix I, is justified by Remark 1 below.

Remark 1. *Suppose that the domain of an attribution rule is enlarged to the class of all TU games and that the five axioms in Theorem 1 are imposed on this enlarged domain, with all quantifiers and relevance interpreted relative to the enlarged domain. Then, for $n \geq 4$, there exists no such attribution rule satisfying all five axioms.*

Proof of Remark 1. Since all games used in the proof of Lemma 1 belong to $\mathbb{V}$, the same argument shows that every coalition in $\mathcal{C}$ is relevant on the unrestricted domain. For $n \geq 4$, information parsimony implies that no coalition outside $\mathcal{C}$ is relevant. Changing the worths of coalitions outside $\mathcal{C}$ one at a time therefore yields the analogue of Corollary 1 on the unrestricted domain. Take $v \notin \mathbb{V}$ with $v(N) \neq v(\emptyset)$, $v(i) = v(\emptyset)$ and $v(N) = v(N \setminus i)$ for every $i \in N$. Applying the construction used in the proof of Claim 5, for each $i \in N$ there exists v^i that agrees with v on $\mathcal{C}$ and in which i is a null player. Hence, $\psi_i(v) = \psi_i(v^i) = 0$ for every $i \in N$, which contradicts efficiency. □

H. Independence of axioms in Theorem 1

In this appendix we demonstrate that each axiom in Theorem 1 is logically independent of the others for $n \geq 4$. To this end, we construct a series of alternative attribution rules, each of which satisfies all axioms except one. These counterexamples show that none of the axioms can be derived from the remaining ones, and hence each plays an essential role in the axiomatic characterization of ψ^{ESENSC}. Establishing such independence strengthens the robustness of the characterization by confirming that the axiomatic system is minimal.

For any $v \in \mathbb{V}$ and any $i \in N$,

- let $\psi_i^1(v) = b_i(v)$,
- let $\psi_i^2(v) = b_i(v) + \frac{1}{n}R(v)$,
- let
$$\psi_i^3(v) = \begin{cases} 0 & \text{if } i \notin \mathcal{A}(v), \\ b_i(v) + \frac{i}{\sum_{j \in \mathcal{A}(v)} j} R(v) & \text{otherwise,} \end{cases}$$
- let
$$\psi_i^4(v) = \begin{cases} 0 & \text{if } i \notin \mathcal{A}(v), \\ \frac{v(N)-v(\emptyset)}{|\mathcal{A}(v)|} & \text{otherwise,} \end{cases}$$
and
- $\psi_i^5(v) = \psi_i^{\mathrm{SHAP}}(v)$.

ψ^1 satisfies all axioms in Theorem 1 but efficiency. ψ^2 satisfies all axioms in Theorem 1 but the null player property. ψ^3 satisfies all axioms in Theorem 1 but restricted differential marginality. ψ^4 satisfies all axioms in Theorem 1 but pairwise interaction splitting. ψ^5 satisfies all axioms in Theorem 1 but information parsimony (for $n \geq 4$).

I. An existing axiomatic characterization of the Shapley value in classical TU games

In this appendix, we briefly review existing axiomatizations of the Shapley value that are closely related to our axiomatization of ESENSC. To relate our framework to the classical cooperative-game-theoretic literature, we restrict attention to the standard domain of TU games in which the value of the empty coalition is normalized to zero: $\bar{\mathbb{V}} = \{\, v : 2^N \to \mathbb{R} \mid v(\emptyset) = 0 \,\}$.[11]

Remark 2 (Corollary 5 in [1])**.** *On $\bar{\mathbb{V}}$, an attribution rule ψ satisfies the following efficiency, the null player property, and differential marginality if and only if $\psi = \psi^{SHAP}$.*

- ***Efficiency.*** *For any $v \in \bar{\mathbb{V}}$, $\sum_{i \in N} \psi_i(v) = v(N) - v(\emptyset)$.*
- ***Null player property.*** *If $i \in N$ satisfies $v(S \cup i) = v(S)$ for any $S \subseteq N \setminus i$, then $\psi_i(v) = 0$.*
- ***Differential marginality.*** *Let $v, w \in \bar{\mathbb{V}}$ and $i, j \in N$ satisfy $v(S \cup i) - v(S \cup j) = w(S \cup i) - w(S \cup j)$ for any $S \subseteq N \setminus \{i, j\}$. Then $\psi_i(v) - \psi_j(v) = \psi_i(w) - \psi_j(w)$.*

It is well known that differential marginality is equivalent to fairness in the sense of van den Brink [17]; see also Proposition 3 in [1]. Moreover, fairness (and therefore differential marginality) can be derived from the combination of two well-known axioms, namely symmetry and additivity (Proposition 2.4 in [17]). Thus, differential marginality is not an entirely independent requirement, but can be understood as a consequence of these more primitive and widely accepted axioms.

[11] The result is considered to remain valid even when $v(\emptyset) \neq 0$. Since the cited paper develops the arguments under the normalization $v(\emptyset) = 0$, we adopt the same convention here for consistency.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction accurately reflect the scope and contributions of the paper. The claims regarding computational efficiency and approximation quality are supported by both the theoretical characterization and empirical evaluation. While the experiments focus on tabular datasets, they are sufficient to support the main claims.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper discusses limitations and scope in the conclusion (Section 6). In particular, it identifies extending ESENSC beyond tabular data as an important direction for future research, highlighting that the current empirical evaluation is limited in scope. It also notes the need to better understand the conditions under which ESENSC and SHAP yield different attribution patterns across model architectures and feature interaction structures. These points clarify the limitations and applicability of the proposed method.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best

judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: All theoretical results are stated with explicit assumptions, and complete formal proofs are provided in the appendix. In the main text, we include intuitive explanations for key results to aid understanding, while maintaining full rigor in the supplementary material.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We provide sufficient details to reproduce the main experimental results in Section 4 and Appendices D and E. Specifically, we describe the datasets used, model architectures, training procedures, and evaluation metrics. Key hyperparameters and implementation choices are also specified. Our method is algorithmically defined in a self-contained manner, allowing independent implementation.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.

(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: The datasets used in our experiments are all publicly available and can be freely downloaded. For baseline methods, we rely on standard implementations (e.g., the SHAP package for exact and approximate SHAP values). The code for our proposed AFA methods is newly implemented by the authors and is currently being prepared for public release. We will release the full implementation, including scripts and instructions to reproduce all experimental results, via a GitHub repository when ready.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: We provide detailed descriptions of the experimental settings, including datasets, data splits, model architectures, training procedures, and evaluation metrics in Section 4. Hyperparameters and their selection procedures are also specified. Additional implementation details are provided in the appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: We do not report statistical significance measures such as error bars or confidence intervals, as our analysis does not focus on statistical inference. Instead, our experiments aim to evaluate feature attribution methods by comparing deterministic outputs (e.g., deviations from SHAP and rank correlations) under fixed experimental settings. The focus of our study is on the decomposition of model predictions rather than on statistical estimation.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: We do not provide detailed information about the compute resources (e.g., hardware specifications or execution time), as our experiments do not require specialized or large-scale computational resources. All experiments can be reproduced using standard computing environments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: The research conducted in this paper conforms to the NeurIPS Code of Ethics. We use publicly available datasets and standard machine learning methods, and no human subjects, sensitive data, or potentially harmful applications are involved.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [No]

Justification:This work is primarily methodological and focuses on feature attribution from a technical perspective. While improved interpretability may have positive societal impacts, and potential risks may arise from misuse or overinterpretation of attribution methods, a detailed discussion of broader societal impacts is beyond the scope of this paper.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: This work does not involve the release of high-risk models or datasets. We use publicly available datasets and propose a feature attribution method, which does not pose significant risks of misuse that would require specific safeguards.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.

- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We use publicly available datasets and standard software packages, and we properly credit their original sources through citations in the paper. The datasets and libraries used (e.g., the SHAP package) are widely used and distributed under standard licenses, and we adhere to their terms of use.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [No]

    Justification: We introduce new code for the proposed AFA methods; however, the corresponding assets are not yet publicly released with full documentation. We are preparing a GitHub repository and will provide well-documented code, including usage instructions and implementation details when ready.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: This work does not involve crowdsourcing or research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: This work does not involve crowdsourcing or research with human subjects, and therefore no IRB approval or equivalent review is required.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core methodology and experimental design of this work do not involve the use of large language models.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.