PADOVA UNIVERSITY PRESS

Dolomites Research Notes on Approximation

Volume 19 · 2026 · Pages 200–216

# Physics-Informed Feature Maps for Kernel-Based Learning: A Reinterpretation of Symbolic Regression

Margherita Lampani[a] · Sabrina Guastavino[a,b] · Michele Piana[a,b] · Federico Benvenuto[a]

*Communicated by Gabriele Santin*

**Abstract**

Supervised machine learning involves approximating an unknown functional relationship from a limited dataset of features and corresponding labels. Within the framework of regularized approximation in reproducing kernel Hilbert spaces (RKHSs), this task reduces to the minimization of a regularized empirical risk functional, where the choice of the kernel encodes prior assumptions on the solution space. In conventional applications, features are standardized and treated as abstract coordinates, disregarding their physical meaning and structural relationships. This abstraction may hinder both interpretability and stability in scientific problems, where dimensional consistency, conservation laws, or symmetries play a crucial role. This study proposes a physics-informed approach to feature-based machine learning that constructs non-linear feature maps informed by physical laws and dimensional analysis. These maps induce physics-informed kernels that embed domain knowledge directly into the geometry of the RKHS, thereby enhancing the interpretability and, when physical laws are unknown, allow for the identification of relevant mechanisms through feature ranking. The method aims to improve both predictive performance in regression tasks and classification skill scores by integrating domain knowledge into the learning process, while also enabling the potential discovery of new physical equations within the context of explainable machine learning.

## 1 Introduction

The problem of learning a functional relationship from finite data can be naturally framed within the theory of regularized approximation in reproducing kernel Hilbert spaces (RKHS) [1]. In that context, the input space $\mathcal{X}$ is a subset of $\mathbb{R}^m$ (or more generally, a measurable space) and given a set of input-output data $\{(x_i, y_i)\}_{i=1}^n \subset \mathcal{X} \times \mathbb{R}$, one seeks a function $g \in \mathcal{H}_K$ minimizing a regularized empirical risk functional of the form

$$\min_{g \in \mathcal{H}_K} \left\{ \sum_{i=1}^n \ell(g(x_i), y_i) + \lambda \|g\|_{\mathcal{H}_K}^2 \right\},$$

where $\ell$ is a suitable loss function [2], $\lambda > 0$ is the regularization parameter [3], and $\mathcal{H}_K$ is a RKHS associated with a positive definite kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ [1,4].

A cornerstone of this framework is the *representer theorem* [5,6], which ensures that the minimizer $g^*$ admits a representation of the form

$$g^*(x) = \sum_{i=1}^n \alpha_i K(x_i, x),$$

thus reducing the infinite-dimensional minimization to a finite-dimensional problem. The structure and properties of the kernel $K$ are therefore central, as they encode prior assumptions on the class of admissible solutions and implicitly define the geometry of the hypothesis space.

In traditional applications, the input space $\mathcal{X}$ is defined in terms of *standardized* or *normalized* features, which are treated as abstract coordinates devoid of semantic or physical interpretation. While such representations are computationally convenient, they neglect important structural information often available in scientific problems–such as dimensional consistency, conservation laws, or symmetries–thereby limiting both the interpretability and the stability of the resulting models. Therefore, in this paper,

[a]MIDA, Dipartimento di Matematica, UniversitÃ  di Genova
[b]Osservatorio Astrofisico di Torino, Istituto Nazionale di Astrofisica

we propose a reinterpretation of the feature construction process inspired by symbolic regression, not with the aim of discovering explicit analytic expressions for the target function $g$, but rather as a mechanism to construct *nonlinear, physics-informed feature maps* $\phi : \Omega \to \mathbb{R}^p$ that respect the physical nature of the problem. Indeed, these maps induce a new kernel

$$K_{\text{phys}}(x, x') = \langle \phi(x), \phi(x') \rangle,$$

which defines a RKHS of functions informed by physical knowledge embedded in the feature transformation $\phi$. The resulting framework enables the design of *physics-informed kernels* that preserve the mathematical structure of kernel-based learning while introducing an additional layer of interpretability grounded in domain knowledge.

Further, this perspective highlights the connection between kernel approximation and inverse problems, since the choice of the feature map $\phi$ (and hence the kernel $K$) can be seen as a form of *prior modeling*. In this sense, the proposed approach can be interpreted as a *physics-informed regularization strategy*, where the regularity is not imposed solely through penalization, but also through the embedding of physical structure in the kernel itself.

This paper aims to formalize this approach to machine learning, and to illustrate its practical utility through several numerical experiments. The methodology is general and applicable to both regression and classification, and we focused on two main scenarios: the first one is when the physical insight is known and can be systematically exploited to enhance both approximation quality and interpretability; the second one is when the physics is unknown. More specifically, in this latter case, we show that the application of feature ranking algorithms to the physics-informed regularization methods allows the selection of the physics-informed features that mostly impact the prediction process, which helps to identify the possible physical mechanism at the base of data interpretation. A posteriori, one should also verify whether, in the case of regression, this approach decreases the error metrics concerned with prediction and, in the case of classification, it is able to improve the confusion matrix associated to the task, thus increasing the corresponding skill scores.

The plan of the paper is as follows. Section 2 offers an overview of related works. Section 3 sets up the mathematical formalism of the physics-informed approach in the context of the connection between kernel approximation and inverse problems. Section 4 describes three applications performed in synthetic settings and related to a standard regression problem, a regression problem where the physics-informed feature corresponding to the corrected physical equation is removed, and a standard classification problem. Section 5 focuses on an experiment leveraging real data concerned with a space weather problem whose physical properties are largely unknown. Our conclusions are offered in Section 6.

## 2 Related works

The incorporation of physical insights into data-driven modeling has received increasing attention in recent years, motivated by the dual goals of enhancing model interpretability and improving generalization. A prominent class of methods addressing this objective is that of *symbolic regression* algorithms, which aim to reconstruct compact analytical expressions governing the observed data. Examples include SINDY (Sparse Identification of Nonlinear Dynamical Systems) [7], AI FEYNMAN [8], and PYSR [9], which combine sparse optimization, neural architectures, or evolutionary heuristics to discover interpretable models reflecting the underlying physics.

These approaches often incorporate physical constraints explicitly, through either dimensional analysis, or conservation laws, or prior structural knowledge. For instance, AI FEYNMAN reduces the search space using invariance properties and dimensional homogeneity, while SINDY can encode known dynamics via a dictionary of admissible terms. Such methodologies are particularly effective when the learning task requires the identification of the governing equations in dynamical systems or of latent physical structures directly from empirical data. From a functional-analytic perspective, these methods implicitly operate within a space of candidate functions shaped by physical priors and aim to minimize some notion of error or misfit, often regularized through sparsity constraints. Nevertheless, they typically treat function discovery as the central goal, making them computationally intensive and less amenable to settings where prediction rather than model identification is the primary objective. It is also important to note their limitations: for example, AI FEYNMAN suffers from a marked sensitivity to noise, which makes it poorly suited for dealing with real experimental data [10].

Unlike symbolic regression, the methodology proposed in the present paper does not aim to yield interpretable symbolic expressions but rather to guide the regularization process through physically meaningful embeddings. This leads to several advantages. First, the approach is computationally more tractable, as it avoids combinatorial search over symbolic expressions. Second, it is compatible with both *regression and classification* tasks, thus extending its applicability beyond traditional symbolic regression frameworks. Third, it can be readily integrated into standard kernel-based learning pipelines, such as ridge regression [11], support vector machines [1], or neural networks [12].

We point out that this perspective aligns closely with the theory of *regularized kernel approximation* [13], where the choice of a kernel implicitly defines a hypothesis space and acts as a form of prior modeling. In this light, the construction of physics-informed feature maps can be seen as a mechanism for embedding domain-specific knowledge into the learning process, thereby functioning as a form of *structured regularization* [14].

Such an approach is particularly valuable in scientific contexts where the physical meaning of features is central to both interpretability and model robustness, and where partial or approximate knowledge of the governing laws is available. Moreover, when the underlying physics is not fully known, feature ranking techniques [15] applied to the constructed feature maps can help identify relevant physical mechanisms *a posteriori*, further supporting model interpretability and potential hypothesis generation.

In summary, while symbolic regression methods are well-suited to the discovery of explicit physical models, the present framework offers a pragmatic alternative for integrating physical reasoning into machine learning tasks through the lens of kernel methods and function space approximation.

## 3 Physics-informed approach to design feature maps and kernels

We introduce our physics-informed approach for kernel-based machine learning by first presenting the connection between approximating a function in a RKHS and solving linear inverse problem where the forward operator $A$ is defined by the physics-informed feature map.

### 3.1 Connection between physics-informed forward operator and Reproducing Kernel Hilbert Spaces

In approximation theory, the function $g$ linking the input features to the output labels is usually assumed to belong to an RKHS, where a reproducing kernel is fixed. This problem is typically addressed by solving a variational problem in which the loss function $\ell$ is arbitrary, and the regularization term generalizes the classical Tikhonov penalty term, depending on the norm $\|\cdot\|_{\mathcal{H}_K}$ in the RKHS $\mathcal{H}_K$. Therefore the following variational problem is considered:

$$\hat{g}_\lambda := \arg\min_{g\in\mathcal{H}_K} \sum_{i=1}^{n} \ell(y_i, g(x_i)) + \lambda\psi(\|g\|_{\mathcal{H}_K}) , \tag{1}$$

where $\lambda > 0$ is the regularization parameter and $\psi : [0, +\infty) \to \mathbb{R}_+$ is a continuous, coercive and strictly monotonically increasing real-valued function. We pointed out that, under the standard assumptions on loss functions, namely that $\ell : \mathbb{R}\times\mathbb{R} \to \mathbb{R}_+$ is such that $\ell(y, \cdot)$ is proper, convex and lower semicontinuous for every $y$, the existence of at least one minimizer of (1) is guaranteed. Moreover, if either $\ell(y, \cdot)$ or $\psi$ is strictly convex, then the minimizer is unique.

A well-established result proves that an RKHS can be viewed as the image space of a compact linear operator $A$ (e.g., see Theorem 4.21 in [16] and analogous versions in [17, 18]). The following proposition is an adaptation of this result in the context of feature maps.

**Proposition 3.1.** *Let us consider a compact linear operator $A : \mathcal{H}_1 \to \mathcal{H}$ where $\mathcal{H}_1$ is a Hilbert space and $\mathcal{H}$ is a Hilbert space of functions over $\Omega$. In this case, for all $x \in \Omega$ there exists an element $\phi(x) \in \mathcal{H}_1$ such that*

$$(Af)(x) = \langle f, \phi(x)\rangle_{\mathcal{H}_1}, \tag{2}$$

*for each element $f \in \mathcal{H}_1$. Then the range of the operator A, denoted as $\Im(A)$, equipped with the norm*

$$\|g\|_{\mathcal{H}_K} = \min\{\|w\|_{\mathcal{H}_1} : w \in \mathcal{H}_1 \ s.t\ g(x) = \langle w, \phi(x)\rangle_{\mathcal{H}_1},\ x \in \Omega\}$$

*is an RKHS with kernel*

$$\begin{aligned} K :\ & \Omega\times\Omega \to \mathbb{R} \\ & (x, r) \to K(x, r) := \langle \phi(x), \phi(r)\rangle_{\mathcal{H}_1}. \end{aligned} \tag{3}$$

We remark that the feature space $\mathcal{F}$ is isometrically isomorphic with $\mathcal{H}_K$. Indeed,

$$\Im(A) = \overline{span\{K(x, \cdot),\ x \in \Omega\}} , \tag{4}$$

and the feature space

$$\mathcal{F} := \overline{span\{\phi(x) ,\ x \in \Omega\}} \tag{5}$$

is such that $\mathcal{F} = \ker(A)^\perp$. Therefore, the first isomorphism theorem applied to the map

$$A : \mathcal{H}_1 \longrightarrow \Im(A) = \mathcal{H}_K \subseteq \mathcal{H} \tag{6}$$

states that

$$\mathcal{F} = \ker(A)^\perp \simeq \mathcal{H}_1/\ker(A) \simeq \Im(A) = \mathcal{H}_K\ . \tag{7}$$

The isomorphism in (7) is induced by the linear operator $A$. More precisely, by defining $\tilde{A} := A_{|\ker(A)^\perp} : \ker(A)^\perp \to \mathcal{H}_K$, is well defined on the quotient space $\mathcal{H}_1/\ker(A)$, since $Af = 0$ for all $f \in \ker(A)$. The map $\tilde{A}$ is bijective. Moreover, for any $f \in \ker(A)^\perp$, we have

$$\|Af\|_{\mathcal{H}_K} = \|f\|_{\mathcal{H}_1},$$

by definition of the norm in $\mathcal{H}_K$, which shows that $A$ is an isometric isomorphism between $\mathcal{F}$ and $\mathcal{H}_K$. Further, we pointed out that the generalized inverse $A^\dagger : \mathfrak{I}(A) + \mathfrak{I}(A)^\perp \to \ker(A)^\perp$ provides the inverse mapping associated with the above isomorphism. Indeed, since $\tilde{A}$ is bijective, its inverse coincides with the restriction of the Moore–Penrose inverse to $\mathcal{H}_K$, namely $\tilde{A}^{-1} = A^\dagger_{|\mathcal{H}_K}$.

Proposition 3.1 and equation (7) allowed us to provide an equivalence result between the approximate solution (1) and the regularized solution

$$\hat{f}_\lambda := \arg\min_{f\in\mathcal{H}_1} \sum_{i=1}^n \ell(y_i, (Af)(x_i)) + \lambda\psi(\|f\|_{\mathcal{H}_1}) \tag{8}$$

of the discretized inverse problem $Af = y$, where $A$ is the operator characterized by the map $\phi$ defined in (2), and $y_i$ for $i = 1, \dots, n$, are noisy samples of the ideal data $y$. Indeed, the following result holds true:

**Proposition 3.2.** *Given $\{(x_i, y_i)\}_{i=1}^n$ a set of samples and by assuming that $K(x, x') = \langle\phi(x), \phi(x')\rangle_{\mathcal{H}_1}$ for all $x, x' \in \Omega$ we have that*

$$\hat{f}_\lambda = A^\dagger \hat{g}_\lambda, \tag{9}$$

*where $A^\dagger : \mathfrak{I}(A) + \mathfrak{I}(A)^\perp \to \ker(A)^\perp$ is the generalized inverse of $A$.*

*Proof.* By hypothesis we have the identification $\mathfrak{I}(A) = \mathcal{H}_K$. Let $\tilde{f} := A^\dagger \hat{g}_\lambda$. By definition of $A^\dagger$, $\tilde{f} \in \ker(A)^\perp$ and therefore, $\|\hat{g}_\lambda\|_{\mathcal{H}_K} = \|\tilde{f}\|_{\mathcal{H}_1}$. For all $f \in \mathcal{H}_1$ we have

$$\begin{aligned} & \sum_{i=1}^n \ell(y_i, (Af)(x_i)) + \lambda\psi(\|f\|_{\mathcal{H}_1}) \\ \geq\ & \min_{g\in\mathfrak{I}(A)} \sum_{i=1}^n \ell(y_i, g(x_i)) + \lambda\psi(\|g\|_{\mathcal{H}_k}) \\ =\ & \sum_{i=1}^n \ell(y_i, \hat{g}_\lambda(x_i)) + \lambda\psi(\|\hat{g}_\lambda\|_{\mathcal{H}_k}) \\ =\ & \sum_{i=1}^n \ell(y_i, A\tilde{f}(x_i)) + \lambda\psi(\|\tilde{f}\|_{\mathcal{H}_1}) \end{aligned} \tag{10}$$

i.e. $\tilde{f}$ is the solution of problem (8). This concludes the proof. □

The connection between $\hat{f}_\lambda$ and $\hat{g}_\lambda$ can be leveraged and interpreted within the context of the proposed physics-informed approach as in Figure 1. This figure explains that, given a set of samples, we can construct a physics-informed feature map by using some knowledge of the feature input (such as the physical meaning or dimensional properties). This can be viewed as defining a linear operator $A$, which encodes the physical model equation, being defined via the physics-informed feature map $\phi$ (see equation (2)). Solving the inverse problem associated with this operator is equivalent to solving a machine learning or approximation problem, where the solution is in a specific Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}_K$, with the kernel $K$ defined by the physics-informed feature map $\phi$, i.e., $K(x, x') = \langle\phi(x), \phi(x')\rangle_{\mathcal{H}_1}$ for all $x, x' \in \mathcal{X}$.

$$\begin{tikzcd}
\{x_i\}_{i=1}^n \arrow[r] & \{\phi(x_i)\}_{i=1}^n \arrow[r] & A \arrow[rr, "\mathcal{H}_K\ =\ \mathfrak{I}(A)", "K(x,x')\ =\ \langle\phi(x),\phi(x')\rangle"'] \arrow[d] & & \mathcal{H}_K \arrow[d] \\
 & \{y_i\}_{i=1}^n \arrow[r] & \hat{f}_\lambda \arrow[rr, "\hat{g}_\lambda = A\hat{f}_\lambda"] & & \hat{g}_\lambda
\end{tikzcd}$$

**Figure 1:** Outline of the process that implements the physics-informed solution of the feature-based machine learning problem: the measured features are transformed by a physics-informed map into an operator $A$ in a Hilbert space setting; the solution $\hat{f}_\lambda$ of the corresponding inverse problem provides the physical equation, which is transformed by $A$ into the predicted output $\hat{g}_\lambda$.

### 3.2 Physics-informed feature map in supervised machine learning

The aim of supervised learning is to find a function $g : \Omega \subset \mathbb{R}^m \longrightarrow \mathbb{R}$ from a set of examples $\{(x_i, y_i)\}_{i=1,\dots,n}$ randomly drawn from an unknown probability distribution $\rho(x, y)$ with $(y_1, \dots, y_n) \in \mathbb{R}^n$ and $(x_1, \dots x_n) \in \Omega^n$, such that $g$ has to explain the relationship between input and output, i.e.

$$y_i \sim g(x_i) \tag{11}$$

for all $i = 1, \dots, n$ and $g(x)$ has to be a good estimate of the output when a new input $x \in \Omega$ is given. A classical setting is given by regression, where it is usually assumed that

$$y_i = g(x_i) + \epsilon_i \ , \tag{12}$$

where $\epsilon_i$ is Gaussian noise, and the sought solution $g$ is a linear function of the form $g(x) = x^T w$, with $w \in \mathbb{R}^m$ being the vector of coefficients to estimate and $x \in \Omega$. This problem is typically studied using variational approaches [19], that is by minimizing the empirical expected value of the loss function over $w \in \mathbb{R}^m$. Since the minimizing solution can be numerically unstable [20], it is common to apply Tikhonov regularization [21], introducing a penalty term to stabilize the solution. The regularized counterpart of the problem is then formulated as the minimization problem

$$\hat{w} = \arg\min_{w \in \mathbb{R}^m} \sum_{i=1}^{n} \ell(y_i, x_i^T w) + \lambda \|w\|_2^2, \tag{13}$$

Using the classical squared loss function leads to ridge regression (also known as penalized least squares) [11], whose explicit solution is given by

$$\hat{w} = (\mathbf{X}^T \mathbf{X} + \lambda I)^{-1} \mathbf{X}^T \mathbf{Y}, \tag{14}$$

where $\mathbf{X}$ represents the $n \times m$ design matrix with $x_i$ as rows, $\mathbf{Y}$ is the $n \times 1$ vector whose components are the features' labels, and $I$ denotes the identity matrix .

However, the relationship between input features and regression output may be more complex than the one modeled by a linear scenario. In such cases, feature maps and kernels provide a way to generalize linear models to non-linear frameworks. In fact, a feature map $\phi : \Omega \to \mathcal{F}$ maps the input space $\Omega$ of the original features to a new feature space $\mathcal{F}$ where a scalar product is defined. This transformation allows the data to be represented in a space where separation performed by means of a linear model is easier. For example, if $\mathcal{F} = \mathbb{R}^p$, the goal is to find a function $g(x_i) = \phi(x_i)^T \beta$, where $\beta \in \mathbb{R}^p$ is the ideal coefficient vector. From a practical perspective, the problem is addressed by solving a similar variational problem, where the ideal coefficient $\beta$ is estimated as

$$\hat{\beta} = (\Phi^T \Phi + \lambda I)^{-1} \Phi^T \mathbf{Y}, \tag{15}$$

where $\Phi$ is the feature map matrix $n \times p$ whose rows are given by a set of feature maps $\{\phi(x_i)\}$. This approach is closely related to kernel methods, since a kernel can be defined through a feature map, i.e. by defining $K(x_i, x_l) = \langle \phi(x_i), \phi(x_l) \rangle_{\mathcal{F}}$.

In the kernel framework, a kernel function (such as a Gaussian or polynomial function) is typically chosen without accounting for the physical meaning or dimensional aspects of the features $x_i$. Our goal was to address this limitation by building new feature maps $\phi$ that combine the original features according to physics-driven laws, resulting in dimensionally homogeneous, physics-informed features (PIFs).

A fundamental requirement of the proposed construction is that all physics-informed features share the same physical dimension as the target variable. This constraint is imposed at the level of the feature map design and ensures that each candidate descriptor is physically consistent with the quantity to be approximated. In this way, dimensional analysis acts as a guiding principle for feature construction, significantly restricting the space of admissible transformations and embedding prior physical knowledge directly into the representation. As a consequence, the learning problem is carried out within a space of dimensionally homogeneous features, which enhances both interpretability and physical coherence of the resulting model.

### 3.3 Physics-Informed Kernel Classification via Soft-Margin SVM

In this section, we formalize the dual formulation of kernel-based classification models when the reproducing kernel is induced by a physics-informed feature map. The structure of the kernel reflects physical constraints or invariants encoded in the feature representation, thereby influencing the geometry of the hypothesis space and the distribution of support vectors.
Let $\phi : \Omega \to \mathcal{H}_1$ be a feature map embedding physical knowledge (e.g., via dimensional consistency or invariants), and let $K(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}_1}$ denote the associated reproducing kernel. Given a set of labeled samples $\{(x_i, y_i)\}_{i=1}^n \subset \mathcal{X} \times \{-1, +1\}$, we consider the primal problem of a soft-margin Support Vector Machine (SVM) in the RKHS $\mathcal{H}_K$:

$$\begin{aligned} \min_{g \in \mathcal{H}_K,\, \xi_i \geq 0,\, b \in \mathbb{R}} \quad & \frac{1}{2} \|g\|_{\mathcal{H}_K}^2 + C \sum_{i=1}^{n} \xi_i \\ \text{subject to} \quad & y_i(g(x_i) + b) \geq 1 - \xi_i, \quad i = 1, \dots, n, \end{aligned} \tag{16}$$

where $C > 0$ is a regularization parameter. The corresponding dual problem admits a representation that explicitly depends on the kernel $K$ as follows.

**Proposition 3.3** (Dual SVM with Physics-Informed Kernel)**.** *Let $\phi : \Omega \to \mathcal{H}_1$ be a physics-informed feature map, and let $K(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}_1}$. Then the dual of the soft-margin SVM problem (16) is*

$$
\begin{aligned}
&\max_{\alpha\in\mathbb{R}^n} \quad && \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j \langle\phi(x_i),\phi(x_j)\rangle_{\mathcal{H}_1} \\
&\textit{subject to} \quad && 0\le\alpha_i\le C, \quad \sum_{i=1}^{n}\alpha_i y_i = 0.
\end{aligned}
\tag{17}
$$

*The resulting classifier $g$ which solves* (16) *is given by*

$$
g(x)=\sum_{i=1}^{n}\alpha_i y_i K(x_i,x)+b=\sum_{i=1}^{n}\alpha_i y_i\langle\phi(x_i),\phi(x)\rangle_{\mathcal{H}_1}+b, \tag{18}
$$

*with $\alpha$ the solution of* (17).

The above result follows from standard kernel SVM theory (see e.g. [1,5]). The structure of the dual solution reveals how physically-informed features influence classification. A point $x_i$ corresponding to a non-zero coefficient $\alpha_i$ is termed *support vector*, and lies on or within the margin. When $\phi(x)$ encodes physically meaningful quantities (e.g., energy, momentum, conserved scalars), the corresponding support vectors are physically critical data points that drive the classification boundary. Even beyond SVM-type algorithms, a physics-informed feature map $\phi(x)$, together with the associated kernel $K(x,x')=\langle\phi(x),\phi(x')\rangle_{\mathcal{H}_1}$, can be seamlessly integrated into a wide range of kernel-based learning algorithms for both regression and classification tasks. Examples include kernel ridge regression (as discussed in section 3.2), kernel Principal Component Analysis (PCA), and even neural networks, where $\phi(x)$ can serve as a physically coherent embedding provided as input to the model.

This general approach introduces an *inductive bias*, a fundamental concept in machine learning that refers to the set of assumptions a learning algorithm relies on to generalize beyond the training data [22–25]. In this context, the inductive bias is explicitly grounded in known physical principles, such as dimensional consistency or conservation laws, which are embedded in the structure of the feature map $\phi$. By incorporating such physically motivated constraints, the resulting models gain interpretability and are better aligned with the underlying physics of the problem domain.

## 4 Experiments in synthetic settings

In this section we show the advantage of constructing physics-informed feature maps in order to define new physics-informed features to infer regularization and classical machine learning methods taking into account the physical dimension. We considered three different theoretical frameworks, i.e. three different model equations and, for each one,

1. We generated an annotated archive of features, where the corresponding labels are computed by solving the model equation and by applying uniform random noise of different intensities. Features for each experiment were obtained by drawing random values within physically plausible intervals, and any derived quantities were subsequently computed from these primary features.
2. We split the annotated archive into a training set and a test set.
3. We applied standardization to generate the corresponding training and test sets of standardized features (SFs).
4. We trained a machine learning algorithm and then applied it to the test set, accordingly computing evaluation metrics such as the mean absolute error (MAE) and the mean squared error (MSE).

Then, we used the same archive of features to generate a training and a test set of PIFs by means of dimensionally homogeneous combinations of the original features, where the dimension of reference is the one of the corresponding labels (see Algorithm 1); the PIFs were then standardized to generate the corresponding set of standardized physics-informed features (SPIFs); the machine learning algorithm optimized by means of the training set of SPIFs was validated by using the corresponding test set; and, finally, evaluation metrics have been computed for comparison with the results obtained using SFs in the training phase. In all experiments considered below, exception made for the application to real data one, the annotated archives are made of 8000 points in both the SFs and PIFs spaces and, in all cases, 70% (6400) of them were used for training and 30% (1600) for testing. We employed Ridge Regression and support vector machine for regression (SVR) to address the regression task, and support vector machine for classification (SVC) to address the classification task.

In order to provide PIFs with a ranking that describes their impact on regression, we applied a recursive feature addition procedure, conceptually related to greedy approaches for feature ranking [26] (see Algorithm 2), where

1. SPIFs are ranked according to the values of the computed regression coefficients.
2. Recursively and starting from the SPIF with highest rank, the prediction is performed up to saturation of MAE and MSE.
3. The regression coefficients of the selected SPIFs are then de-standardized in order to determine the (physically meaningful) regression coefficients associated to the corresponding PIFs.

This last step allows the most significant physical equation used to generate the PIFs to be identified. In the following we report two remarks regarding concerning the feature-ranking procedure and the determination of the coefficients associated with the SPIFs.

*Remark* 1. We point out that the standardization/de-standardization process of the PIFs reduces possible numerical issues in the computation of the weights while allowing the determination of the physical coefficients associated to the model equations.

*Remark* 2. We highlight that the stopping criterion adopted in Algorithm 2 is intentionally based on predictive performance rather than on the variation between consecutive predictors. Specifically, the inclusion of a new physics-informed feature is evaluated through the corresponding improvement in the approximation of the target variable, as measured by a global error metric. This choice is consistent with the philosophy of greedy feature selection methods, where the relevance of a feature is assessed according to the gain in predictive performance achieved by the learning algorithm. In particular, our approach is inspired by classifier-dependent greedy feature selection strategies [26], where the stopping rule is determined by the saturation of a performance score rather than by the magnitude of the changes in the model coefficients or predictions. From this perspective, the objective of Algorithm 2 is not to quantify the sensitivity of the predictor to the addition of a feature, but rather to identify the point beyond which additional physics-informed features do not provide a significant improvement in generalization performance.

**Algorithm 1** PIF generation

**Input**: $x \in \mathbb{R}^m$ feature vector for each sample; $[y]$ physical dimension of the label.
**Step 1**: Generate the list of features' dimensions
**Step 2**: Define $p$ rules to combine features, i.e. define a feature map $\phi : \mathbb{R}^m \to \mathbb{R}^p$ such that $[\phi(x)_l] = [y]$ for $l = 1, \dots, p$
**Output**: $\text{PIF}_l = \phi(x)_l$ for $l = 1, \dots, p$

The three experiments presented in the following subsections each highlight a different aspect of the PIF-based approach. In the first one, concerning fluid dynamics, the Bernoulli equation is expressed as a linear combination of three distinct PIFs, each corresponding to one of its physical terms; the goal is therefore to assess whether the algorithm can both select the correct subset of PIFs and recover their respective coefficients. Moreover, this first experiment was also used to compare our feature ranking algorithm (Algorithm 2) with an already established symbolic regression method, namely PySR [27]. In the second one, concerning magnetic energy dissipation in pulsars, we investigate how the algorithm's behaviour changes depending on whether the exact governing equation is included among the input PIFs or not. In the third experiment, we assess the validity of the PIF methodology in a classification setting, and we further investigate the impact of class imbalance by systematically varying the proportions of the two classes in the dataset.

The code used to reproduce the results associated with this manuscript is published on GitHub (`https://github.com/margherita-lampani/Physics-Informed-Feature-Maps-for-Kernel-Based-Learning`).

### 4.1 First example: fluid dynamics

Bernoulli equation [28] represents the fundamental law that models the dynamics of the ideal fluid under stationary flow conditions. It is expressed as

$$p + \frac{\rho v^2}{2} + \rho g h = \text{const}, \tag{21}$$

where $\rho$ is the fluid density, $v$ is its speed, $p$ denotes the static pressure of the fluid (either water or air in this experiment), and $\frac{\rho v^2}{2}$ is the dynamic pressure arising from the fluid's motion; finally, $\rho g h$ corresponds to the hydrostatic pressure exerted by a column of fluid, where $g$ is the gravitational acceleration and $h$ is the height of the fluid column. Together with the aforementioned features, we incorporated additional features essential for the study of a fluid dynamic system, i.e., the volumetric flow rate $Q$, the cross-sectional area of the column $A$, and the dynamic viscosity $\mu$, all sampled independently from uniform distributions over plausible physical ranges: velocity $v \in [0.01, 100]$ m/s, volumetric flow rate $Q \in [0.001, 10]$ m$^3$/s, cross-sectional area $A \in [0.01, 5]$ m$^2$, height $h \in [0.01, 100]$ m. Since water and air are considered as fluid types with equal probability, $\rho$, $\mu$ and $p$ assume different values accordingly: for water ($\rho = 1000$ kg/m$^3$, $\mu = 10^{-3}$ Pa · s and $p \in [10^3, 10^7]$ Pa) while for air ($\rho = 1.225$ kg/m$^3$, $\mu = 1.8 \times 10^{-5}$ Pa · s and $p \in [10^2, 10^5]$ Pa). We combined these seven features to generate seven PIFs characterized by the dimension of a pressure, since this is the physical dimension of the quantity being predicted, where the first three PIFs are the ones incorporated in the Bernoulli equation (21). From a mathematical viewpoint, in this example $x = (p, \rho, v, Q, A, \mu, h)$ and we defined the physics-informed feature map $\phi : \mathbb{R}^7 \to \mathbb{R}^7$ such as

$$\phi(x) = (p, \rho v^2, \rho g h, \frac{pQ}{Av}, \frac{\mu v}{h}, \frac{\mu Q}{hA^2}, \frac{\rho A g}{h}). \tag{22}$$

Table 1 contains the definitions of all original features and PIFs together with the corresponding units.

For this first experiment, we also include a comparison with a state-of-the-art symbolic regression method, PySR [27]. PySR is a multi-population evolutionary algorithm with asynchronous evolution, widely used for discovering interpretable equations from data, and it allows enforcing the physical units of the target quantity.

**Algorithm 2** Feature ranking

**Input**: $\hat{\beta}$ estimated coefficient vector defined in (15); $\tilde{\phi}(x_i)$ SPIF, i.e. standardized PIF, for each sample $x_i$ for $i = 1, \dots, n$; $y_i$ standardized label associated to each sample for $i = 1, \dots, n$; $\tau > 0$ tolerance for the stop criterion.
**Define** the set of the ordered indexes $J$ with respect to the absolute value of the components of $\hat{\beta}$, i.e. $|\hat{\beta}_{J(1)}| \geq |\hat{\beta}_{J(2)}| \geq |\hat{\beta}_{J(p)}|$.
**For** $k = 1, \dots, p$

- **Make** the prediction using the first $k$ relevant PIFs, i.e. for $i = 1, \dots, n$ compute

$$\hat{y}_i^{(k)} = \sum_{j=1}^{k} \tilde{\phi}(x_i)_{J(j)} \hat{\beta}_{J(j)}. \tag{19}$$

- **Compute** the mean absolute error between the predicted and actual labels, i.e.

$$\mathrm{MAE}^{(k)} = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - \hat{y}_i^{(k)} \right| \tag{20}$$

**if** $k > 1$ and $|\mathrm{MAE}^{(k)} - \mathrm{MAE}^{(k-1)}| < \tau$ **then** stop
**Output**: $J(1), \dots, J(k)$ indexes of the selected PIFs.

| Feature | Unit | PIF | Unit |
|---|---|---|---|
| $F_1 = p$ | $\frac{kg}{m \cdot s^2}$ | $\mathrm{PIF}_1 = p$ | Pa |
| $F_2 = \rho$ | $\frac{kg}{m^3}$ | $\mathrm{PIF}_2 = \rho v^2$ | Pa |
| $F_3 = v$ | $\frac{m}{s}$ | $\mathrm{PIF}_3 = \rho g h$ | Pa |
| $F_4 = Q$ | $\frac{m^3}{s}$ | $\mathrm{PIF}_4 = \frac{pQ}{Av}$ | Pa |
| $F_5 = A$ | $m^2$ | $\mathrm{PIF}_5 = \frac{\mu v}{h}$ | Pa |
| $F_6 = \mu$ | $\frac{kg}{m \cdot s}$ | $\mathrm{PIF}_6 = \frac{\mu Q}{hA^2}$ | Pa |
| $F_7 = h$ | $m$ | $\mathrm{PIF}_7 = \frac{\rho A g}{h}$ | Pa |

**Table 1:** Prediction of Bernoulli equation. Features and PIFs generated by a feature map triggered by the pressure dimension. $\mathrm{PIF}_1$, $\mathrm{PIF}_2$, and $\mathrm{PIF}_3$ are the PIFs that generate equation (21).

The results of the application of the regression algorithms to both SFs and SPIFs are given in Table 2 and Figure 2. Table 2 reports the Mean Absolute Errors (MAEs) and Mean Squared Errors (MSEs) obtained with Ridge Regression, SVR, and PySR. Figure 2 represents Ridge Regression box-plots where, for better visualization, we included only the inter-quartile range (IQR box-plots from now on). As the results provided by SVR are very similar, we did not include them in the figure.

We then applied the feature ranking algorithm to SPIFs and showed in Figure 3 that MAE and MSE saturation occurs starting from $\mathrm{SPIF}_4$. The de-standardization of the coefficients associated to $\mathrm{SPIF}_1$, $\mathrm{SPIF}_2$, and $\mathrm{SPIF}_3$ leads to the physics-informed regression coefficients in Table 3. These coefficients for the two methods are clearly an accurate approximation of the physical coefficients in the Bernoulli equation (21). We remark that the purpose of constructing multiple physics-informed features is not to assume that all of them contribute significantly to the prediction task. Rather, the proposed framework generates a set of dimensionally consistent and physically plausible candidate descriptors, from which the ranking procedure identifies the most informative ones. Therefore, the fact that several PIFs exhibit a negligible impact on the prediction is an expected outcome and it demonstrates the ability of the ranking strategy to isolate the physically relevant mechanisms among a broader set of admissible candidates. This aspect is particularly important in applications where the governing physical law is only partially known, since the relevant combinations cannot be identified a priori and must be inferred from data.

In Table 4 we report the equations obtained by each method across different noise levels. As shown, our algorithm consistently recovers the correct physical expression, with coefficients very close to the true ones. In contrast, PySR never selects the correct Bernoulli form as the best model according to its internal scoring system, which balances predictive accuracy and model complexity. Although the correct expression appears among the candidate formulas at the lowest noise level, it is not chosen as the final model. As the noise level increases, the physically correct expression is no longer identified among the candidate equations.

| Method | | Noise level 10% MAE | MSE | 30% MAE | MSE | 50% MAE | MSE |
|---|---|---|---|---|---|---|---|
| **Ridge Regression** | SFs | 0.179 | 0.052 | 0.275 | 0.168 | 0.353 | 0.330 |
| | SPIFs | 0.067 | 0.017 | 0.189 | 0.135 | 0.281 | 0.301 |
| **SVR (linear kernel)** | SFs | 0.167 | 0.055 | 0.252 | 0.176 | 0.329 | 0.338 |
| | SPIFs | 0.068 | 0.017 | 0.193 | 0.135 | 0.282 | 0.302 |
| **PySR** | | 0.086 | 0.024 | 0.198 | 0.144 | 0.326 | 0.377 |

**Table 2:** MAE and MSE computed from the results provided by ridge regression and SVR with linear kernel when SFs and SPIFs are used to predict the Bernoulli equation (21). The table also reports the results obtained with the symbolic regression method PySR, which is included as a benchmark comparison.

Three different levels of uniform noise were applied to the synthetic labels.

| | Noise level 10% | 30% | 50% | |
|---|---|---|---|---|
| | Estimated coefficient | | | Ground-truth coefficient |
| Method | **Ridge Regression** | | | |
| $SPIF_1$ | 0.998 | 0.993 | 0.990 | 1 |
| $SPIF_2$ | 0.500 | 0.501 | 0.502 | 0.5 |
| $SPIF_3$ | 1.005 | 1.024 | 1.031 | 1 |
| Method | **SVR (linear kernel)** | | | |
| $SPIF_1$ | 0.996 | 0.987 | 0.988 | 1 |
| $SPIF_2$ | 0.498 | 0.490 | 0.491 | 0.5 |
| $SPIF_3$ | 1.033 | 1.037 | 0.960 | 1 |

**Table 3:** Prediction of the Bernoulli equation. Best ranked PIFs and corresponding regression coefficients for three levels of uniform noise affecting the labels.

| | Noise level 10% | 30% | 50% |
|---|---|---|---|
| Method | **Ridge Regression - SPIFs** | | |
| Equation | $(0.998 \cdot p) + (0.500 \cdot \rho v^2) + (1.005 \cdot \rho g h)$ | $(0.993 \cdot p) + (0.501 \cdot \rho v^2) + (1.024 \cdot \rho g h)$ | $(0.990 \cdot p) + (0.502 \cdot \rho v^2) + (1.031 \cdot \rho g h)$ |
| Method | **SVR (linear kernel) - SPIFs** | | |
| Equation | $(0.996 \cdot p) + (0.498 \cdot \rho v^2) + (1.033 \cdot \rho g h)$ | $(0.987 \cdot p) + (0.490 \cdot \rho v^2) + (1.037 \cdot \rho g h)$ | $(0.988 \cdot p) + (0.491 \cdot \rho v^2) + (0.960 \cdot \rho g h)$ |
| Method | **PySR** | | |
| Equation | $p + (\rho \cdot 44.799) \cdot v$ | $\frac{p}{0.759}$ | $p \cdot 1.318$ |

**Table 4:** Equations selected by the feature ranking algorithm (Algorithm 2) for Ridge Regression and SVR with a linear kernel across the different noise levels. For PySR, the reported equations are those achieving the highest score according to the method's internal ranking.

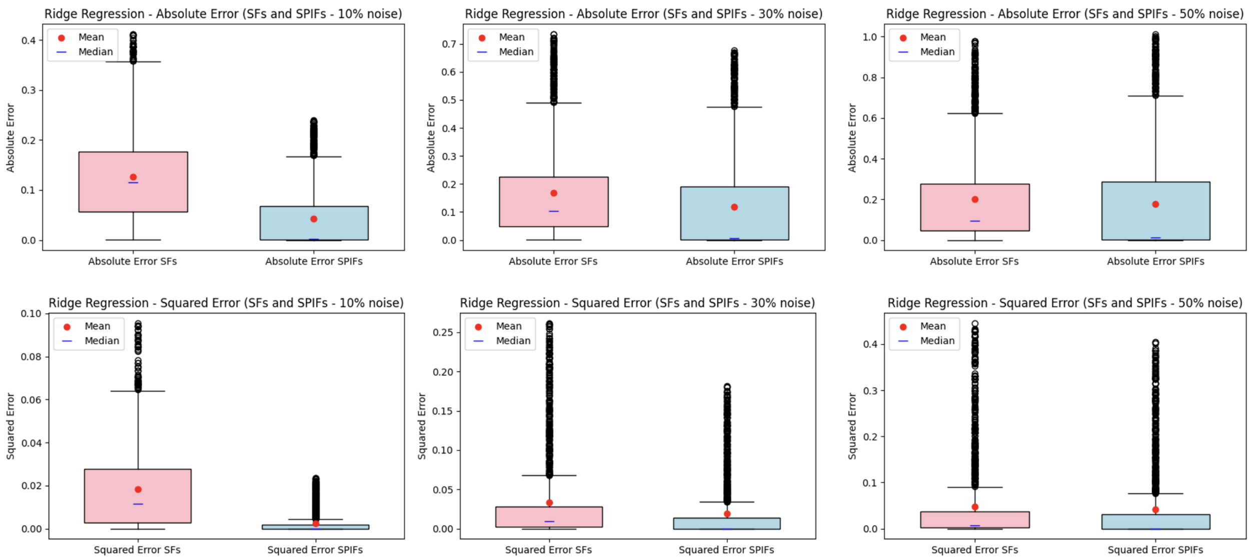


**Figure 2:** Boxplots of the IQR distributions for absolute errors (top panel) and squared errors (bottom panel) provided by ridge regression for predicting the Bernoulli equation, by varying the noise levels. The pink and light blue boxlots represent the results obtained using SFs and SPIFs, respectively.

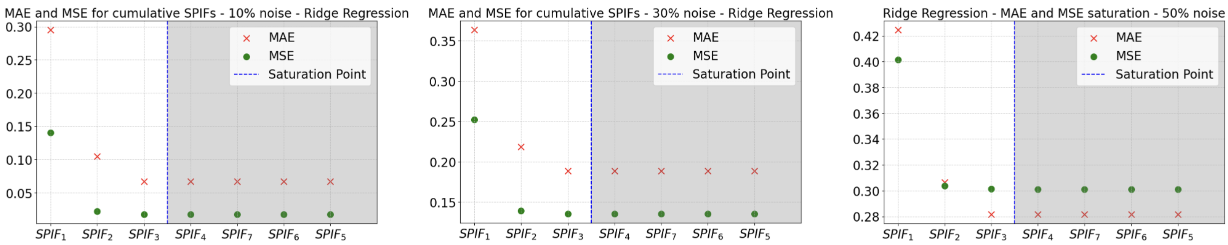


**Figure 3:** Prediction of the Bernoulli equation. Saturation of MAE and MSE while applying the sequential feature ranking Algorithm 2.

### 4.2 Second example: magnetic dissipation in energetic pulsars

Pulsars are highly magnetized rotating neutron stars that emit beams of electromagnetic radiation [29]. Through these emissions, they lose magnetic energy according to the dissipation law [30]

$$\frac{dE}{dt} = -\frac{2\pi B^2 r^6 \omega^4 \sin^2\alpha}{3\mu_0 c^3}, \tag{23}$$

where $B$ is the magnetic field of the star, $r$ is its radius, $\omega$ is its angular velocity, $\alpha$ is its inclination angle, $\mu_0$ is the magnetic permeability in vacuum and $c$ is the speed of light in vacuum. In addition to these quantities, we introduced four more representative features of the system: $P = \frac{2\pi}{\omega}$, the rotational period of the star; $m$, its mass; $I = \frac{2mr^2}{5}$, the moment of inertia; and $E = \frac{I\omega^2}{2}$, the rotational kinetic energy. All features are sampled independently from uniform distributions over plausible astrophysical ranges: radius $r \in [10^4, 10^6]$ m, surface magnetic field $B \in [10^8, 10^{12}]$ T, angular velocity $\omega \in [1, 10^3]$ rad/s, inclination angle $\alpha \in [0, \pi/2]$ rad, and mass $m \in [2.8 \times 10^{30}, 4 \times 10^{30}]$ kg (corresponding to 1.4–2 $M_\odot$, where $M_\odot$ denotes the solar mass); the remaining features $P$, $I$, and $E$ are derived deterministically from these. As usual, PIFs were defined according to the physical unit of the label, i.e., power, and, again, $\text{PIF}_1$ represents the exact equation for the magnetic energy dissipation as in (23) but excluding the dimensionless coefficients. Both features and PIFs are summarized in Table 5.

$\text{PIF}_1$ in Table 5 is tagged by a $*$ for the following reason. In order to study the robustness of the PIF-based approach with respect to the dimensionally-coherent combination of features, we have run the experiment twice, where, in the first case, all SPIFs were used to train the algorithm; in the second run, however, we have excluded $\text{SPIF}_1$ from the training phase (note that for both runs all SFs have been always included). The box-plots in Figure 4 shows that, for both runs of the experiment, training machine learning with SPIFs led to significantly more accurate predictions than when SFs are utilized. However, the errors significantly increase when $\text{SPIF}_1$, which corresponds to equation (23) describing the actual physical process, is excluded from

| Feature | Unit | PIF | Unit |
|---|---|---|---|
| $F_1 = r$ | $m$ | $PIF_1^* = -\frac{B^2 r^6 \omega^4 sin^2(\alpha)}{\mu_0 c^3}$ | W |
| $F_2 = B$ | $T$ | $PIF_2 = -\frac{B^2 c^4 sin^2(\alpha)}{\omega^3 \mu_0} v^2$ | W |
| $F_3 = \omega$ | $\frac{1}{s}$ | $PIF_3 = -\frac{B^2 r^4 \omega}{\mu_0}$ | W |
| $F_4 = \alpha$ | rad | $PIF_4 = -E/P$ | W |
| $F_5 = P$ | $S$ | $PIF_5 = -I/P^3$ | W |
| $F_6 = m$ | $kg$ | $PIF_6 = -mr^2\omega^3$ | W |
| $F_7 = I$ | $kg \cdot m^2$ | $PIF_7 = -\frac{mr^2}{\omega^3}$ | W |
| $F_8 = E$ | $\frac{kg \cdot m^2}{s^2}$ | | |

**Table 5:** Prediction of the pulsar equation (23). Features and PIFs with respective units. The $*$ for $PIF_1$, which corresponds to the correct physical equation, denotes the fact that the second run of the experiment has been carried out without that PIF.

the training, while still remaining lower than those of the classical SF approach. The error depends on how closely the PIFs resemble the correct equation. Again, we chose to report only the plots for Ridge Regression, as the results from SVR exhibited a very similar behavior.

### 4.3 Third example: binary system

We then studied whether our PIF-based approach to machine learning could also be reliably applied to a classification problem. Specifically, in this third experiment our aim was to determine whether or not a pair of stars or celestial bodies were gravitationally bound, having at disposal measures of the energies of the binary system, the bound energy being defined as [31]

$$E = \frac{1}{2} \frac{m_1 \cdot m_2}{(m_1 + m_2)} \cdot v^2 - G \frac{m_1 \cdot m_2}{r}, \tag{24}$$

where $m_1$ and $m_2$ represent the masses of the bodies, $v$ is their relative velocity, $r$ is their distance, and $G$ is the gravitational constant. The dataset consists of 8000 synthetic binary systems, with masses $m_1, m_2 \in [0.5, 10]\ M_\odot$, orbital separations $r \in [0.1, 1000]$ AU, and relative velocities sampled conditionally on the class: bound systems ($E < 0$, label 1) have $v \in [0, v_{\text{crit}})$, while unbound systems ($E > 0$, label 0) have $v \in [v_{\text{crit}}, 2\,v_{\text{crit}})$, where $v_{\text{crit}} = \sqrt{2G(m_1 + m_2)/r}$ is the escape velocity. In order to study the effect of class imbalance, we generated three configurations with controlled proportions of bound to unbound systems: a balanced dataset (50/50), a moderately imbalanced dataset (70/30), and a strongly imbalanced dataset (90/10). Features and PIFs considered for this experiment are listed in Table 6.

| Feature | Unit | PIF | Unit |
|---|---|---|---|
| $F_1 = m_1$ | kg | $PIF_1 = \frac{m_1 m_2}{m_1 + m_2} v^2$ | J |
| $F_2 = m_2$ | kg | $PIF_2 = -G\frac{m_1 m_2}{r}$ | J |
| $F_3 = v$ | m/s | $PIF_3 = \frac{m_1^2}{m_2} v^2$ | J |
| $F_4 = r$ | m | $PIF_4 = -G\frac{m_1^2}{r}$ | J |
| | | $PIF_5 = \frac{m_2^2}{m_1} v^2$ | J |
| | | $PIF_6 = -G\frac{m_2^2}{r}$ | J |

**Table 6:** Classification of binary systems. Features and PIFs with respective units.

The application of SVC with both linear and gaussian kernels to the test sets of SFs and SPIFs provided the confusion matrices in Table 7. In each one of these matrices entry $(1,1)$ contains the true negatives (TNs), entry $(2,2)$ contains the true positives (TPs), entry $(1,2)$ contains the false positives (FPs), and entry $(2,1)$ contains the false negatives (FNs).

From these confusion matrices we computed several skill scores, always reported in Table 7, assessing the classification effectiveness of the algorithm, i.e., the Sensitivity

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{25}$$

| | Dataset percentages of bound and unbound events | | | | | |
|---|---|---|---|---|---|---|
| | 50/50 | | 70/30 | | 90/10 | |
| Method | **SVC with linear kernel** | | | | | |
| Metric | SFs | SPIFs | SFs | SPIFs | SFs | SPIFs |
| TSS | 0.763 | 0.988 | 0.744 | 0.997 | 0.347 | 0.998 |
| HSS | 0.762 | 0.988 | 0.709 | 0.997 | 0.135 | 0.990 |
| Sensitivity | 0.851 | 0.991 | 0.868 | 0.999 | 0.588 | 0.998 |
| Specificity | 0.912 | 0.996 | 0.875 | 0.998 | 0.759 | 0.994 |
| Accuracy | 0.881 | 0.994 | 0.871 | 0.999 | 0.605 | 0.998 |
| Confusion matrix | $\begin{bmatrix} 726 & 70 \\ 120 & 684 \end{bmatrix}$ | $\begin{bmatrix} 793 & 3 \\ 7 & 797 \end{bmatrix}$ | $\begin{bmatrix} 428 & 61 \\ 146 & 965 \end{bmatrix}$ | $\begin{bmatrix} 488 & 1 \\ 1 & 1110 \end{bmatrix}$ | $\begin{bmatrix} 120 & 38 \\ 594 & 848 \end{bmatrix}$ | $\begin{bmatrix} 158 & 0 \\ 3 & 1439 \end{bmatrix}$ |
| Method | **SVC with gaussian kernel** | | | | | |
| Metric | SFs | SPIFs | SFs | SPIFs | SFs | SPIFs |
| TSS | 0.956 | 0.974 | 0.960 | 0.970 | 0.977 | 0.981 |
| HSS | 0.956 | 0.974 | 0.956 | 0.968 | 0.942 | 0.962 |
| Sensitivity | 0.972 | 0.989 | 0.982 | 0.988 | 0.989 | 0.994 |
| Specificity | 0.984 | 0.985 | 0.978 | 0.982 | 0.987 | 0.987 |
| Accuracy | 0.978 | 0.987 | 0.981 | 0.986 | 0.989 | 0.993 |
| Confusion matrix | $\begin{bmatrix} 783 & 13 \\ 22 & 782 \end{bmatrix}$ | $\begin{bmatrix} 784 & 12 \\ 9 & 795 \end{bmatrix}$ | $\begin{bmatrix} 478 & 11 \\ 19 & 1092 \end{bmatrix}$ | $\begin{bmatrix} 480 & 9 \\ 13 & 1098 \end{bmatrix}$ | $\begin{bmatrix} 156 & 2 \\ 15 & 1427 \end{bmatrix}$ | $\begin{bmatrix} 156 & 2 \\ 9 & 1433 \end{bmatrix}$ |

**Table 7:** Classification of binary systems. Skill scores using SVC with both linear and gaussian kernels.

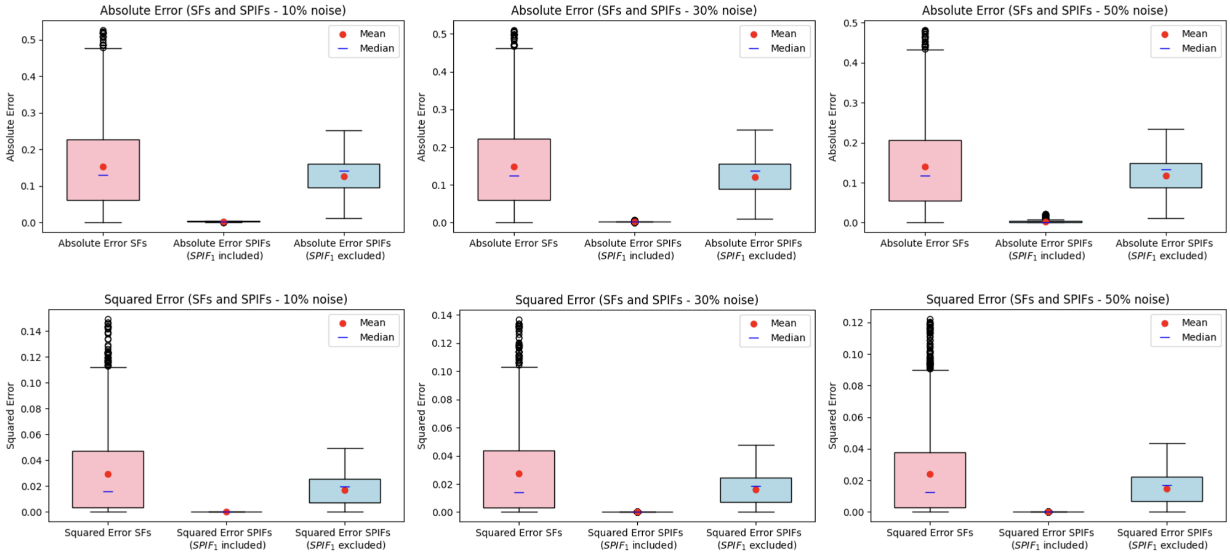


**Figure 4:** Prediction of the pulsar equation. In each panel, from left to right, the IQR box-plots correspond to the result provided by ridge regression when the training is performed by using all SFs, all SPIFs and all SPIFs except $\mathrm{SPIF}_1$, respectively. The absolute errors and squared errors by varying the noise level are represented in the top and bottom panels, respectively.

the Specificity

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \,, \tag{26}$$

the Accuracy

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \,, \tag{27}$$

the True Skill Statistic (TSS)

$$\text{TSS} = \text{Sensitivity} + \text{Specificity} - 1 \,, \tag{28}$$

and the Heidke Skill Score (HSS)

$$\text{HSS} = \frac{2 \cdot (\text{TP} \cdot \text{TN} - \text{FN} \cdot \text{FP})}{(\text{TP} + \text{FN}) \cdot (\text{FN} + \text{TN}) + (\text{TP} + \text{FP}) \cdot (\text{FP} + \text{TN})} . \tag{29}$$

We highlight that when the linear kernel is used, the SVC classifier operates directly in the SPIF space and the kernel coincides with the inner product induced by the physics-informed feature map, namely $K(x, x') = \langle \phi(x), \phi(x') \rangle$, as described in Proposition 3.3. Therefore, the resulting decision boundary is linear in the SPIF representation and can be interpreted as a classifier acting on physically meaningful quantities. We pointed out that Gaussian kernel is applied to the SPIF representation, i.e. the radial basis function is applied to $z = \phi(x)$ and $z' = \phi(x')$, $K_{\text{RBF}}(z, z') = \exp\left(-\gamma \|z - z'\|^2\right)$, with $\gamma > 0$ is the kernel parameter. Therefore, the nonlinear kernel acts on a feature space that already incorporates physically meaningful combinations and dimensional constraints. Therefore, the Gaussian kernel introduces an additional nonlinear transformation of the SPIF space, allowing us to assess separately the effect of the physics-informed representation and that of a more expressive kernel.

## 5 Application to real data: solar flare forecasting

Solar flares are the most energetic events in the heliosphere [32]. Their significance is two-fold. On the one hand, the physical mechanisms at their base are still mostly unknown and the identification of the model equations that realistically describe them is a crucial open issue in solar physics [33]. On the other hand, solar flares are the main trigger of space weather, so that the realization of a reliable flare forecasting process may have a crucial impact on the safeguard of several both in space and on Earth assets [34, 35].

The use of machine learning for flare forecasting is rather recent, and typically relies on the well-established result that these explosions are generated by active regions on the solar atmosphere characterized by a complex magnetic configuration. The flare forecasting game leveraging feature-based supervised machine learning approaches is therefore typically based on the following process [36–41]:

| Feature | Unit | PIF | Unit |
|---|---|---|---|
| $F_1 = I$ | $A$ | $PIF_1 = \rho \cdot l \cdot S$ | $T \cdot A \cdot m^2$ |
| $F_2 = F$ | $T \cdot A \cdot m$ | $PIF_2 = \Phi \cdot I$ | $T \cdot A \cdot m^2$ |
| $F_3 = H$ | $\frac{T^2}{m}$ | $PIF_3 = \frac{F \cdot B^2}{H}$ | $T \cdot A \cdot m^2$ |
| $F_4 = \Phi$ | $T \cdot m^2$ | $PIF_4 = \frac{H \cdot I^2 \cdot S^2}{F}$ | $T \cdot A \cdot m^2$ |
| $F_5 = S$ | $m^2$ | $PIF_5 = \frac{H \cdot I \cdot S}{\nabla B}$ | $T \cdot A \cdot m^2$ |
| $F_6 = \rho$ | $\frac{T \cdot A}{m}$ | $PIF_6 = \nabla B \cdot l \cdot I \cdot S$ | $T \cdot A \cdot m^2$ |
| $F_7 = B$ | $T$ | | |
| $F_8 = \nabla B$ | $\frac{T}{m}$ | | |
| $F_9 = l$ | $m$ | | |

**Table 8:** Solar flare forecasting. Features and PIFs with respective units.

1. The magnetograms in the database of the Helioseismic and Magnetic Imager on board the NASA Solar Dynamics Observatory [42], which represents the most complete and up-to-date archive of active regions images, are annotated by means of the light-curves measured by the Geostationary Operational Environmental Satellites (GOES), which reveal the flare presence by means of a continuous monitoring of the X-ray emission from the Sun.
2. A feature extraction algorithm is applied to the magnetograms to extract large sets of active regions descriptors.
3. A machine learning algorithm is trained by using the Helioseismic and Magnetic Imager (HMI) features and the corresponding GOES binary labels.

In this experiment we applied this process by using the nine features indicated in the first column of Table 8, which are considered among the most predictive descriptors of solar flares. These selected features represent various physical quantities, including electric current ($I$), force ($F$), magnetic helicity ($H$), magnetic flux ($\Phi$), area ($S$), energy density ($\rho$), magnetic field strength ($B$), magnetic field gradient ($\nabla B$), and characteristic length ($l$). These features have been extracted from the portion of the HMI archive in the time interval between 9/15/2012 and 09/07/2017. From these features we have generated the PIFs in the second column of the same table, where we chose energy as the triggering dimension (we note that GOES provides also the class of the flare, which is related to the energy it releases during its explosive phase). We address the task as a binary classification problem, where the goal is to predict the occurrence of at least one intense solar flare (above the M GOES class) within the next 24 hours. In this case, we assigned label 1 to all observations exhibiting at least one flare, and label 0 to those in which no flare was detected. The annotated archive made of active region magnetograms has been chronologically split into two sets made of 70% and 30% of the whole database, respectively. In this experiment, samples associated with active regions that produce C-class flares in the next 24 hours are discarded in order to make a clearer distinction between positive and negative samples. SVCs with linear and gaussian kernels were trained using the first portion of the corresponding SFs and SPIFs, and tested against the second one, to obtain the confusion matrices from which we computed the skill scores described in Table 9. From the results, we observed a significant improvement in all skill scores when SPIFs were used in both cases. Furthermore, we employed a ranking approach similar to the one presented in Algorithm 2, where the estimated weights from the SVC with a linear kernel were used to rank the PIFs. We then progressively added one PIF at a time to observe the behaviour of the TSS. We pointed out that the estimated coefficients by SVC define the direction of the separating hyperplane and provide an indication of the relative relevance of each PIF, although they can not be directly interpreted as absolute measures of importance. In detail, when the linear kernel is used, simple computations from the equation (17) leads to the following representation $g(x) = \sum_{i=1}^{n} \alpha_i y_i \langle \phi(x), \phi(x_i) \rangle_{\mathbb{R}^p} + b = \phi(x)^T \beta + b$, where $\beta = \sum_{i=1}^{n} \alpha_i y_i \phi(x_i)$, which can be linked to the formula in (19). We pointed out, that in this case, for each step, the SVC was re-trained using the corresponding subset of 1, 2, and so on PIFs. In Figure 5 the results are shown: in this case, the saturation point is defined as the point at which a decrease in the score is observed and no further improvement occurs beyond the maximum value obtained when $PIF_2$ is added.

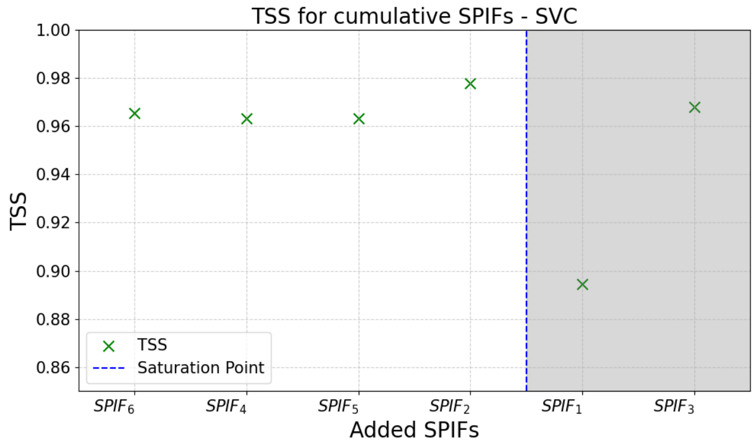


**Figure 5:** Solar flare forecasting experiment. TSS obtained using an SVC with an analogous sequential feature ranking algorithm (see Algorithm 2). In this experiment, the SVC is re-trained each time a PIF is added, and the saturation point is defined as the point at which a decrease in the score is observed and no further improvement occurs beyond the maximum value that is obtained.

| | **SVC (linear)** | | **SVC (gaussian)** | |
|---|---|---|---|---|
| Metric | SFs | SPIFs | SFs | SPIFs |
| TSS | 0.835 | 0.968 | 0.856 | 0.982 |
| HSS | 0.229 | 0.447 | 0.297 | 0.587 |
| Sensitivity | 0.909 | 1. | 0.909 | 1. |
| Specificity | 0.926 | 0.968 | 0.947 | 0.982 |
| Accuracy | 0.926 | 0.969 | 0.947 | 0.982 |
| Confusion matrix | $\begin{bmatrix} 755 & 60 \\ 1 & 10 \end{bmatrix}$ | $\begin{bmatrix} 789 & 26 \\ 0 & 11 \end{bmatrix}$ | $\begin{bmatrix} 772 & 43 \\ 1 & 10 \end{bmatrix}$ | $\begin{bmatrix} 800 & 15 \\ 0 & 11 \end{bmatrix}$ |

**Table 9:** Solar flare forecasting. Skill scores computed from confusion matrices using SVC with linear and gaussian kernels.

## 6 Conclusions

This study introduces a physics-informed feature-based approach to machine learning. In this approach, feature maps are designed to generate physics-informed features characterized by the same physical dimension. On the one hand, experiments with synthetic data showed that 1) training machine learning with these PIFs enhances both the regression and the classification performances of the algorithm; and 2) the constructed regression coefficients provide accurate estimate of the equation describing the data formation process.

An experiment on flare forecasting realized by means of real data showed that the improvement in the classification power still persists. However, a simplified setting was considered to better distinguish between positive and negative classes; therefore, further investigations are needed since the equations used to generate the PIFs were designed without full knowledge of the physics underlying flare mechanisms, which remain largely unknown. Interestingly, the ranking algorithm applied to the PIFs in this study highlights the most predictive PIFs, among which is $\text{PIF}_2 = \Phi I$. In fact, [43] provides a physical justification for the $\text{PIF}_2$ quantity in the framework of current-driven flare and Coronal Mass Ejections (CMEs) models, where it can be interpreted as the released energy during a flare, taking into account the electromotive force, the associated current, and the internally transported energy during the process. This may pave the way for a new research line, where the systematic generation of PIFs could be utilized to identify unknown descriptors of how energy is distributed in active regions as precursors of flaring activity.

Finally, beyond the specific applications considered in this work, the proposed approach may also be interpreted from a broader mathematical perspective as a strategy for the construction of approximation spaces informed by physical principles. The physics-informed feature map induces a kernel and, consequently, a reproducing kernel Hilbert space whose geometry reflects prior physical knowledge encoded through dimensional consistency and physically motivated feature design. From this viewpoint, the methodology shifts the emphasis from the design of the predictor to the design of the function space in which approximation is performed. This perspective highlights a potential connection between machine learning, kernel methods, and approximation theory, where physics-based feature maps act as a mechanism for defining structured and interpretable hypothesis spaces.

At the same time, a limitation of the proposed framework is that, in problems involving a large number of input variables, the manual generation of all possible physics-informed features may become impractical due to the rapid growth of the number of candidate combinations. However, we emphasize that the proposed methodology is not intended to replace automatic feature-selection or feature-learning approaches, but rather to complement them. By restricting the search space to physically meaningful and dimensionally consistent representations, physics-informed features provide a principled way to incorporate prior scientific knowledge into the learning process while reducing the complexity associated with unconstrained feature generation. Future research will investigate the integration of the proposed framework with automatic feature-selection techniques and feature-learning architectures, with the aim of extending its applicability to high-dimensional and more complex scenarios.

## Acknowledgments

ML and MP acknowledge the support of the Fondazione Compagnia di San Paolo within the framework of the Artificial Intelligence Call for Proposals AIxtreme project (ID Rol: 71708). MP also acknowledges the financial support of the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 104 published on 2.2.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union - NextGenerationEU - Project Title "Inverse Problems in Imaging Sciences (IPIS)" - CUP D53D23005740006 - Grant Assignment Decree No. 973 adopted on 30/06/2023 by the Italian Ministry of Ministry of University and Research (MUR). FB acknowledges financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No.1409 published on 14.9.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union - NextGenerationEU - Project Title "COronal mass ejection, solar eNERgetic particle and flare forecaSTing from phOtospheric sigNaturEs (CORNERSTONE)" - CUPD53D23022570001 - Grant Assignment Decree No. 1294 adopted on 4.8.2023 by the Italian Ministry of Ministry of University and Research (MUR). All authors are also grateful to the Gruppo Nazionale per il Calcolo Scientifico - Istituto Nazionale di Alta Matematica (GNCS - INdAM). This research has been performed in the framework of the MIUR Excellence Department Project awarded to Dipartimento di Matematica, Università di Genova, CUP D33C23001110001.

### References

[1] Vladimir Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.

[2] Lorenzo Rosasco, Ernesto De Vito, Andrea Caponnetto, Michele Piana, and Alessandro Verri. Are loss functions all the same? *Neural computation*, 16(5):1063–1076, 2004.

[3] Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[4] Vern I Paulsen and Mrinal Raghupathi. *An introduction to the theory of reproducing kernel Hilbert spaces*, volume 152. Cambridge university press, 2016.

[5] Bernhard Schölkopf, Ralf Herbrich, and Alex J Smola. A generalized representer theorem. In *International conference on computational learning theory*, pages 416–426. Springer, 2001.

[6] Ernesto De Vito, Lorenzo Rosasco, Andrea Caponnetto, Michele Piana, and Alessandro Verri. Some properties of regularized kernel methods. *The Journal of Machine Learning Research*, 5:1363–1390, 2004.

[7] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

[8] Silviu-Marian Udrescu and Max Tegmark. Ai feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16):eaay2631, 2020.

[9] Miles Cranmer, Alvaro Sanchez-Gonzalez, Peter Battaglia, Rui Xu, Kyle Cranmer, David Spergel, and Shirley Ho. Discovering symbolic models from deep learning with inductive biases, 2020.

[10] Liron Simon Keren, Alex Liberzon, and Teddy Lazebnik. A computational framework for physics-informed symbolic regression with straightforward integration of domain knowledge. *Scientific Reports*, 13(1):1249, 2023.

[11] Gary C McDonald. Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1):93–100, 2009.

[12] Sun-Chong Wang. Artificial neural network. In *Interdisciplinary computing in java programming*, pages 81–100. Springer, 2003.

[13] Charles A Micchelli, Massimiliano Pontil, and Peter Bartlett. Learning the kernel function via regularization. *Journal of machine learning research*, 6(7), 2005.

[14] Xu Sun. Structure regularization for structured prediction. *Advances in Neural Information Processing Systems*, 27, 2014.

[15] Yin-Wen Chang and Chih-Jen Lin. Feature ranking using linear svm. In *Causation and prediction challenge*, pages 53–64. PMLR, 2008.

[16] Ingo Steinwart and Andreas Christmann. *Support vector machines*. Springer Science & Business Media, 2008.

[17] Rainer Kress. *Linear integral equations*, volume 82. Springer, 1989.

[18] Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3):337–404, 1950.

[19] Cheng Zhang, Judith Bütepage, Hedvig Kjellström, and Stephan Mandt. Advances in variational inference. *IEEE transactions on pattern analysis and machine intelligence*, 41(8):2008–2026, 2018.

[20] Tomaso Poggio and Vincent Torre. Ill-posed problems and regularizaron analysis in early vision. *Artificial Intelligence Lab. Memo*, (773), 1984.

[21] Gene H Golub, Per Christian Hansen, and Dianne P O'Leary. Tikhonov regularization and total least squares. *SIAM journal on matrix analysis and applications*, 21(1):185–194, 1999.

[22] Jonathan Baxter. A model of inductive bias learning. *Journal of artificial intelligence research*, 12:149–198, 2000.

[23] Ning Miao. *Incorporating inductive biases into machine learning algorithms*. PhD thesis, University of Oxford, 2024.

[24] Tom M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.

[25] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[26] Fabiana Camattari, Sabrina Guastavino, Francesco Marchetti, Michele Piana, and Emma Perracchione. Classifier-dependent feature selection via greedy methods. *Statistics and Computing*, 34(5):151, 2024.

[27] Miles Cranmer. Interpretable machine learning for science with pysr and symbolicregression. jl. *arXiv preprint arXiv:2305.01582*, 2023.

[28] John David Anderson and John Wendt. *Computational fluid dynamics*, volume 206. Springer, 1995.

[29] Fridolin Weber. *Pulsars as astrophysical laboratories for nuclear and particle physics*. Routledge, 2017.

[30] Hayk Hakobyan, Alexander Philippov, and Anatoly Spitkovsky. Magnetic energy dissipation and $\gamma$-ray emission in energetic pulsars. *The Astrophysical Journal*, 943(2):105, 2023.

[31] Ronald W Hilditch. *An introduction to close binary stars*. Cambridge University Press, 2001.

[32] Michele Piana, A Gordon Emslie, Anna Maria Massone, and Brian R Dennis. *Hard X-ray Imaging of Solar Flares*, volume 164. Springer, 2022.

[33] Einar Tandberg-Hanssen and A Gordon Emslie. *The physics of solar flares*, volume 14. Cambridge University Press, 1988.

[34] Enrico Camporeale, Simon Wing, and Jay Johnson. *Machine learning techniques for space weather*. Elsevier, 2018.

[35] Manolis K. Georgoulis, Stephanie L. Yardley, Jordan A. Guerra, Sophie A. Murray, Azim Ahmadzadeh, Anastasios Anastasiadis, Rafal Angryk, Berkay Aydin, Dipankar Banerjee, Graham Barnes, Alessandro Bemporad, Federico Benvenuto, D. Shaun Bloomfield, Monica Bobra, Cristina Campi, Enrico Camporeale, Craig E. DeForest, A. Gordon Emslie, David Falconer, Li Feng, Weiqun Gan, Lucie M. Green, Sabrina Guastavino, Mike Hapgood, Dustin Kempton, Irina Kitiashvili, Ioannis Kontogiannis, Marianna B. Korsos, K.D. Leka, Paolo Massa, Anna Maria Massone, Dibyendu Nandy, Alexander Nindos, Athanasios Papaioannou, Sung-Hong Park, Spiros Patsourakos, Michele Piana, Nour E. Rawafi, Viacheslav M. Sadykov, Shin Toriumi, Angelos Vourlidas, Haimin Wang, Jason T. L. Wang, Kathryn Whitman, Yihua Yan, and Andrei N. Zhukov. Prediction of solar energetic events impacting space weather conditions. *Advances in Space Research*, 2024.

[36] Sabrina Guastavino, Francesco Marchetti, Federico Benvenuto, Cristina Campi, and Michele Piana. Operational solar flare forecasting via video-based deep learning. *Frontiers in Astronomy and Space Sciences*, 9:1039805, 2023.

[37] Sabrina Guastavino, Francesco Marchetti, Federico Benvenuto, Cristina Campi, and Michele Piana. Implementation paradigm for supervised flare forecasting studies: A deep learning application with video data. *Astronomy & Astrophysics*, 662:A105, 2022.

[38] Domenico Cicogna, Francesco Berrilli, Daniele Calchetti, Dario Del Moro, Luca Giovannelli, Federico Benvenuto, Cristina Campi, Sabrina Guastavino, and Michele Piana. Flare-forecasting algorithms based on high-gradient polarity inversion lines in active regions. *The Astrophysical Journal*, 915(1):38, 2021.

[39] Manolis K Georgoulis, D Shaun Bloomfield, Michele Piana, Anna Maria Massone, Marco Soldati, Peter T Gallagher, Etienne Pariat, Nicole Vilmer, Eric Buchlin, Frederic Baudin, et al. The flare likelihood and region eruption forecasting (flarecast) project: flare forecasting in the big data & machine learning era. *Journal of Space Weather and Space Climate*, 11:39, 2021.

[40] Cristina Campi, Federico Benvenuto, Anna Maria Massone, D Shaun Bloomfield, Manolis K Georgoulis, and Michele Piana. Feature ranking of active region source properties in solar flare forecasting and the uncompromised stochasticity of flare occurrence. *The Astrophysical Journal*, 883(2):150, 2019.

[41] Kostas Florios, Ioannis Kontogiannis, Sung-Hong Park, Jordan A Guerra, Federico Benvenuto, D Shaun Bloomfield, and Manolis K Georgoulis. Forecasting solar flares using magnetogram-based predictors and machine learning. *Solar Physics*, 293(2):28, 2018.

[42] Jørgen Schou, Philip H Scherrer, Rock Irvin Bush, Richard Wachter, S'ebastien Couvidat, Maria Cristina Rabello-Soares, Richard S Bogart, JT Hoeksema, Y Liu, TL Duvall, et al. Design and ground calibration of the helioseismic and magnetic imager (hmi) instrument on the solar dynamics observatory (sdo). *Solar Physics*, 275:229–259, 2012.

[43] DB Melrose. Current-driven flare and cme models. *Journal of Geophysical Research: Space Physics*, 122(8):7963–7978, 2017.