metals

MDPI

Article

Prediction of the Transition-Temperature Shift Using Machine Learning Algorithms and the Plotter Database

Diego Ferreño [1,*], Marta Serrano [2], Mark Kirk [3] and José A. Sainz-Aja [1]

[1] Laboratory of Science and Engineering of Materials Division (LADICIM), University of Cantabria, E.T.S. de Ingenieros de Caminos, Canales y Puertos, Av. Los Castros 44, 39005 Santander, Spain; jose.sainz-aja@unican.es
[2] Division of Energy Interest Materials, CIEMAT, Avda. Complutense, 40, 28040 Madrid, Spain; marta.serrano@ciemat.es
[3] Central Research Institute of Electric Power Industry, Yokosuka 40-0196, Japan; kirk@peaiconsulting.com
* Correspondence: ferrenod@unican.es

Abstract: The long-term operating strategy of nuclear plants must ensure the integrity of the vessel, which is subjected to neutron irradiation, causing its embrittlement over time. Embrittlement trend curves used to predict the dependence of the Charpy transition-temperature shift, ΔT_{41J}, with neutron fluence, such as the one adopted in ASTM E900-15, are empirical or semi-empirical formulas based on parameters that characterize irradiation conditions (neutron fluence, flux and temperature), the chemical composition of the steel (copper, nickel, phosphorus and manganese), and the product type (plates, forgings, welds, or so-called standard reference materials (SRMs)). The ASTM (American Society for Testing and Materials) E900-15 trend curve was obtained as a combination of physical and phenomenological models with free parameters fitted using the available surveillance data from nuclear power plants. These data, collected to support ASTM's E900 effort, open the way to an alternative, purely data-driven approach using machine learning algorithms. In this study, the ASTM PLOTTER database that was used to inform the ASTM E900-15 fit has been employed to train and validate a number of machine learning regression models (multilinear, k-nearest neighbors, decision trees, support vector machines, random forest, AdaBoost, gradient boosting, XGB, and multi-layer perceptron). Optimal results were obtained with gradient boosting, which provided a value of R^2 = 0.91 and a root mean squared error ≈10.5 °C for the test dataset. These results outperform the prediction ability of existing trend curves, including ASTM E900-15, reducing the prediction uncertainty by ≈20%. In addition, impurity-based and permutation-based feature importance algorithms were used to identify the variables that most influence ΔT_{41J} (copper, fluence, nickel and temperature, in this order), and individual conditional expectation and interaction plots were used to estimate the specific influence of each of the features.

Keywords: machine learning; neutron embrittlement; gradient boosting

Citation: Ferreño, D.; Serrano, M.; Kirk, M.; Sainz-Aja, J.A. Prediction of the Transition-Temperature Shift Using Machine Learning Algorithms and the Plotter Database. *Metals* **2022**, *12*, 186. https://doi.org/10.3390/met12020186

Academic Editor: Ferenc Gillemot

Received: 16 December 2021
Accepted: 17 January 2022
Published: 19 January 2022

1. Introduction

Nuclear reactor pressure vessel (RPV) steels are degraded by neutron irradiation. This leads to an increase in strength and a decrease in toughness, which produces the ductile-to-brittle transition temperature shift (TTS), ΔT_{41J} that occurs over the operating lifetime of the plant (T_{41J} represents the temperature at which the energy absorbed in a Charpy test is 41J; this value is widely considered as a reliable definition for the location of the ductile-to-brittle transition region). The long-term operation (LTO) strategy for a nuclear plant must ensure the integrity of the vessel. To determine the TTS in advance, nuclear utilities conduct surveillance programs that place capsules holding specimens fabricated with the same steel as that of the vessel in the beltline region of the vessel. Thus, the specimen irradiation history mimics the neutron spectrum, temperature history, and maximum neutron fluence experienced by the reactor vessel inner surface. Low surveillance capsule lead factors

(i.e., the ratio of the neutron fluence rate, E > 1 MeV, at the specimens in a surveillance capsule to the neutron fluence rate, E > 1 MeV, at the reactor pressure vessel inside the surface peak fluence location) of 3× to 5× support the assumption that the irradiation response of the specimens in the capsules is representative of the steel in the vessel wall. The objectives of a reactor vessel surveillance program are twofold: to monitor changes in the fracture toughness properties and to make use of the data obtained to determine the conditions under which the vessel can be operated throughout its service life. For historical reasons [1], the Charpy test was selected in surveillance programs to monitor the irradiation-induced degradation of the materials. More specifically, the parameter used in most countries to measure the degree of embrittlement is the TTS at 41J absorbed energy.

The prediction of neutron irradiation embrittlement of RPV steels is of utmost importance for demonstrating the integrity of light water reactors. A number of embrittlement trend curves (ETC) have been developed for this purpose [2–5]. Some of these [2,3,5] addressed trends in national data sets and, as such, used smaller empirical data sets as part of their calibration, However, in 2015 the ASTM subcommittee on Nuclear Structural Materials approved ASTM Standard E900-15 [4] to predict the radiation-induced TTS in reactor vessel materials based on an extensive database (nearly 1900 TTS measurements) of surveillance information. The TTS proposed embrittlement correlation was developed using the following variables: copper, nickel, phosphorus and manganese contents, irradiation temperature, neutron fluence as well as the product type (forgings, plates and SRM plates, and welds). Only power reactor (pressurized water reactors, PWR, and boiling water reactors, BWR) surveillance data were used to calibrate this ETC. While E900-15 is an empirical fit, it nevertheless adopts an equation form common to mechanistically guided ETCs having two additive terms: one depending on copper and one independent of copper.

A significant part of the ASTM effort in developing E900-15 focused on the collection, curation, and verification of a large international dataset on embrittlement quantified using both TTS and yield strength increase (ΔYS). These data came from surveillance reports on light water reactors (LWR) and from the technical literature. Attention focused on steels of the type used in LWRs of western design; ex-Soviet water-water energetic reactor (VVER) steels were not considered. The final database (called "PLOTTER") included 4438 TTS or ΔYS data records (36% from PWR surveillance programs, 8% from BWR surveillance programs, and 56% from material test reactor (MTR) research programs obtained at neutron fluxes generally exceeding those common in PWRs). From these, a "BASELINE" data subset was defined to develop what became the TTS equation in E900-15. The BASELINE subset included steels of commercial grade having all known descriptive variables (copper, nickel, manganese, phosphorus, fluence, flux, temperature, and product form). These steels were exposed to neutron irradiation in a PWR or BWR, and had embrittlement quantified by the TTS measured using full-size Charpy V-notch specimens. The BASELINE subset included 1878 TTS surveillance data from 13 countries: Brazil, Belgium, France, Germany, Italy, Japan, Mexico, The Netherlands, South Korea, Sweden, Switzerland, Taiwan, and the United States.

The classical 'model-driven' paradigm in scientific research has been complemented over the last decades with the so-called 'data-driven' approach [6]. Machine learning (ML) methods enable the solution of problems that otherwise could not be addressed by standard statistical/analytical fitting approaches. Rather than designing a complex experiment or developing a numerical model, it is possible using ML to extract patterns from large ensembles of data that are often heterogeneous and/or incomplete. These techniques are progressively gaining use in scientific fields such as materials science [7–10].

This paper was aimed at obtaining a regression ML model that, using the information provided by the PLOTTER database, enables prediction of the TTS. Prediction is important, but understanding is of no less importance. For this reason, specific algorithms have been implemented after fine-tuning the ML models to identify the variables that most influence the TTS as well as to quantify their individual influence and the interaction between variables. The remainder of the paper is organized as follows: the PLOTTER database, the

ASTM E900-15 ETC [4] and the ML methods are described in Section 2. Section 3 is devoted to presenting the results of the analysis. Finally, the interpretation and significance of the results are discussed in Section 4.

2. Methods

2.1. The ASTM PLOTTER Database

The dataset employed to train the ML models corresponds to the BASELINE subset from the PLOTTER database (described in Section 1), which contains 1878 observations. In this data collection:

- The TTS is the target (or variable to be predicted).
- The predictor (or regressor) variable includes numeric vales to describe the chemical composition (Cu, Ni, P, Mn) and irradiation conditions (neutron fluence, flux and temperature) and also indicator/categorical variables describing the product type (welds, plates, forgings, or SRM plates) and the reactor type (BWR or PWR).

Figure 1 shows the histograms of the numeric features, including the target response, the TTS (Figure 1a), while the two categorical attributes are represented as barplots in Figure 2. As can be seen, the database contains a larger proportion of PWR than BWR vessels. Additionally, as made clear by both the plots of Figures 1 and 2, the data structure is highly non-uniform; data are very dense in some regions and sparse (or non-existent) in others. These characteristics of the data structure occur as a consequence of both the specifications to which the RPV steels were procured as well as the operating and design characteristics of both BWR (Boiling Water Reactors) and PWR (Pressurized Water Reactors) reactors.

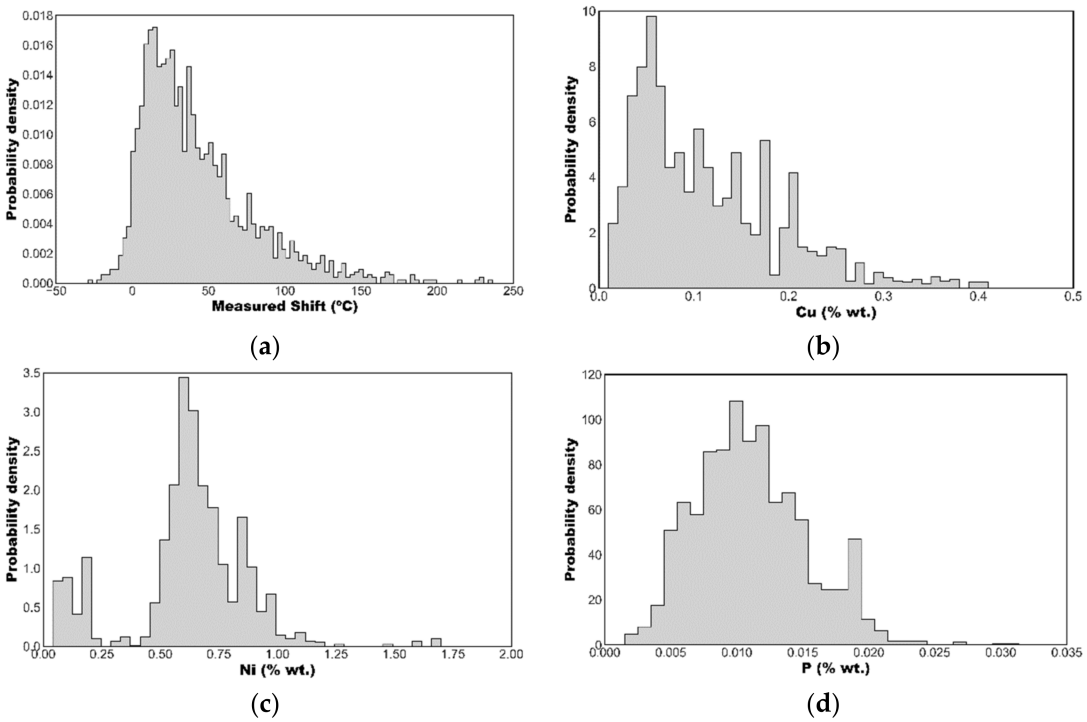


Figure 1. *Cont.*

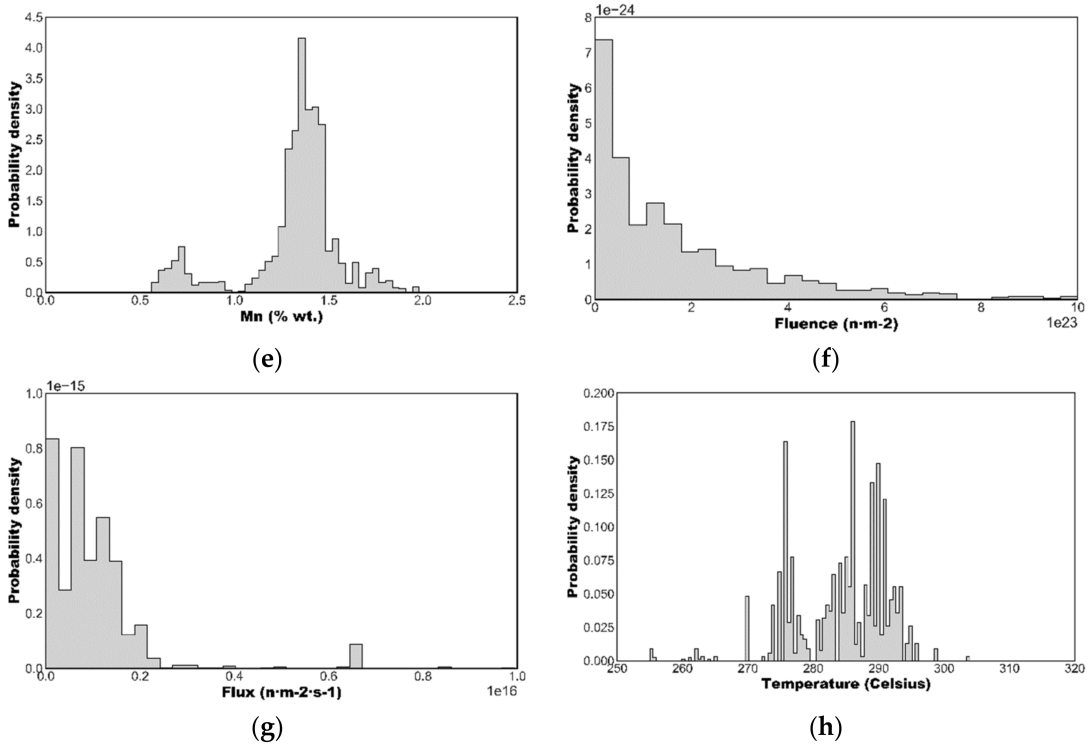


Figure 1. Histograms showing the distribution of the numeric features in the PLOTTER database. (**a**) TTS, (**b**) copper, (**c**) nickel, (**d**) phosphorus, (**e**) manganese, (**f**) fluence, (**g**) flux and (**h**) temperature.

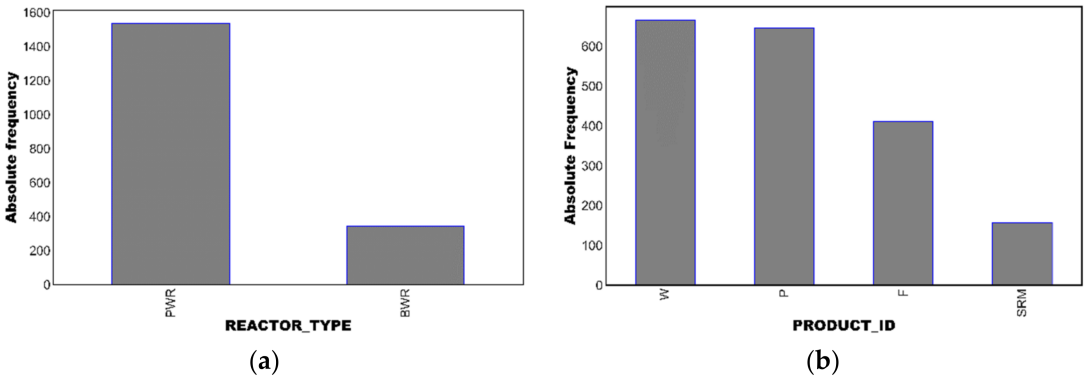


Figure 2. Barplots showing the distribution of the two nominal features in the PLOTTER database. (**a**) Reactor type (PWR: pressurized water reactors or BWR: boiling water reactors) and (**b**) product type (welds, plates, forgings and SRMs: standard reference materials).

2.2. The ASTM E900-15 Embrittlement Trend Curve

The ASTM effort to develop what became E900-15 [4] determined the following analytical expression for TTS in ASTM E900-15 using the BASELINE dataset from the PLOTTER database. The mean value of TTS, in °C, is calculated as follows (1)–(6):

$$TTS = TTS_1 + TTS_2 \quad (1)$$

$$TTS_1 = A\frac{5}{9}1.8943 \cdot 10^{-12}\Phi^{0.5695}\left(\frac{1.8T+32}{550}\right)^{-5.47}$$
$$\left(0.09+\frac{P}{0.012}\right)^{0.216}\left(1.66+\frac{Ni^{8.54}}{0.63}\right)^{0.39}\left(\frac{Mn}{1.36}\right)^{0.3} \quad (2)$$

$$A = \begin{pmatrix} 1.011\ for\ forgings \\ 1.080\ for\ plates\ and\ SRM\ plates \\ 0.919\ for\ welds \end{pmatrix} \quad (3)$$

$$TTS_2 = \frac{5}{9}\ max[min(Cu,\ 0.28) - 0.053,\ 0]\ M \quad (4)$$

$$M = B\ max\left\{min\left[113.87\ \left(ln(\Phi) - ln\left(4.5{\cdot}10^{20}\right)\right),\ 612.6\right],\ 0\right\}\left(\frac{1.8\ T + 32}{550}\right)^{-5.45}$$
$$\left(0.1 + \frac{P}{0.012}\right)^{-0.098}\left(0.168 + \frac{Ni^{0.58}}{0.63}\right)^{0.73} \quad (5)$$

$$B = \begin{pmatrix} 0.738\ for\ forgings \\ 0.819\ for\ plates\ and\ SRM\ plates \\ 0.968\ for\ welds \end{pmatrix} \quad (6)$$

In this equation, Cu, Ni, P and Mn are all expressed in weight percent, Φ is in n/m^2 (E > 1 MeV), and T is in °C. This analytical model includes 26 free parameters that were fitted using a maximum likelihood procedure from the data in the BASELINE dataset. Relative to the calibration dataset, the model provides unbiased predictions and has a root mean square error (RMSE) value of 13.3 °C. E900-15 adopts the following Formulation (7) for standard deviation, which increases along with the predicted value of TTS:

$$SD = \begin{bmatrix} W: 7.681 \\ P: 6.593 \\ F:\ 6.972 \end{bmatrix} \times TTS^{\left[\begin{array}{ccc} W: 0.181 & P: 0.163 & F: 0.199 \end{array}\right]} \quad (7)$$

2.3. Machine Learning

The ML models have been developed and evaluated in the Python 3 programming language using libraries such as Numpy, Pandas, Scikit-Learn, Matplotlib and Seaborn, among others. The workflow of this ML project is summarized [11,12] in the following sections.

2.3.1. Scope of the Analysis

A regression analysis is aimed at predicting a numeric value for new input data. Here, the target variable (that is, the variable to be predicted) is the TTS, ΔT_{41J}. The predictors correspond to the nine features included in the PLOTTER database (copper, nickel, phosphorus, manganese, fluence, flux, temperature, product type and reactor type). The dataset included 1878 instances.

2.3.2. Data Preprocessing

The ability to learn from ML models and the useful information that can be derived may be extremely influenced by data preprocessing. This consists of cleaning the raw data to enable the optimization of the model. Preprocessing includes the following stages [11,12]:

- Data outliers can result in longer training times and less accurate models. Outliers, which were defined as data points beyond a *z*-score (see Formula (8) for the definition of the z-score of the observation x_i, where m_x is the sample mean and s_x is the sample standard deviation) $|z| > 3.0$, were not observed in the PLOTTER BASELINE dataset.

$$z_i = \frac{x_i - m_x}{s_x} \quad (8)$$

- Multicollinearity is potentially harmful for the performance of the model; it may reduce its statistical significance and make it difficult to determine the importance of a feature to the target variable. The Pearson's correlation matrix (see Expression (9) for the definition of the sample Pearson correlation coefficient, *r*) of the dataset, Figure 3, was estimated to identify correlations between features. It was decided to remove one

of the features of every couple with a correlation coefficient exceeding (in absolute value) 0.60. Nevertheless, since the maximum correlations observed were between Cu and P (r = 0.48), no features were eliminated.

$$r = \frac{\sum_{i=1}^{n}(x_i - m_x)(y_i - m_y)}{\sqrt{\sum_{i=1}^{n}(x_i - m_x)^2}\sqrt{\sum_{i=1}^{n}(y_i - m_y)^2}} \quad (9)$$

- Standardization/feature scaling of a dataset is mandatory for some ML algorithms and advisable for others. In this study, features were scaled through the StandardScaler provided by Scikit-Learn [13] which standardizes the features by removing the mean and scaling to unit variance.
- Nominal categorical variables (reactor type and product id) were subjected to the Scikit-Learn OneHotEncoder [13].

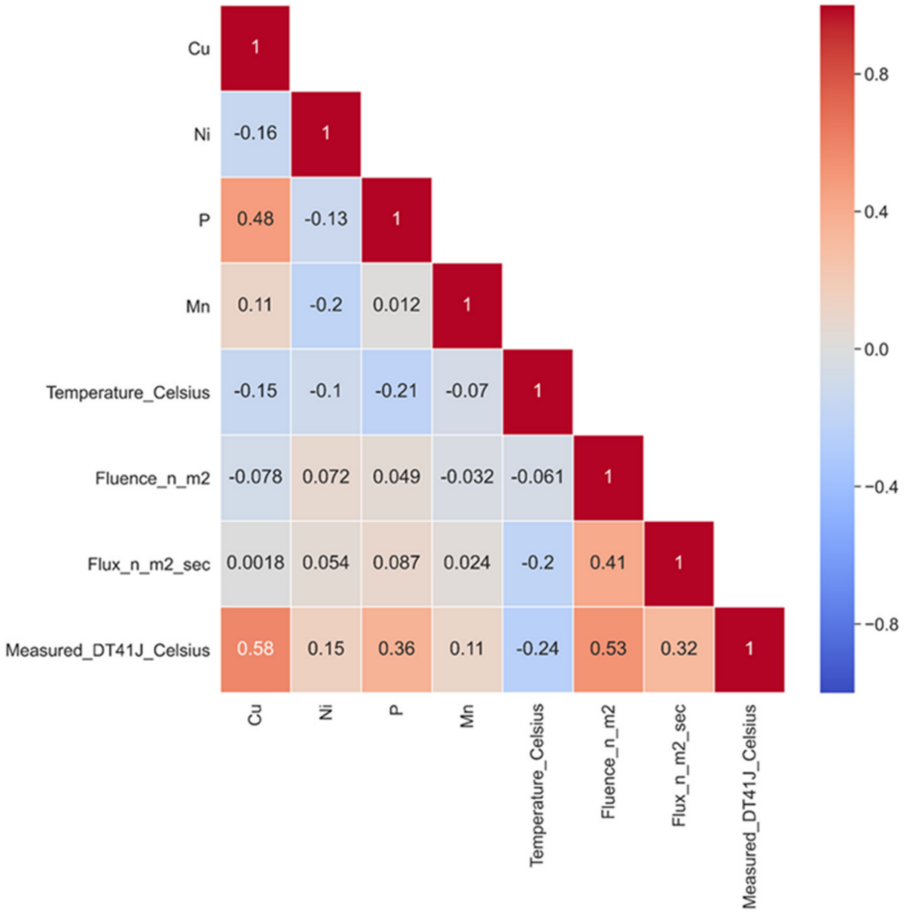


Figure 3. Pearson's correlation matrix of the dataset.

2.3.3. ML Algorithms

As stated by Wolpert in his "No Free Lunch theorem" of ML, "[. . .] for any two learning algorithms, there are just as many situations (appropriately weighted) in which algorithm one is superior to algorithm two as vice versa, according to any of the measures of superiority" [14]. In addition, "if an algorithm does particularly well on average for one class of problems then it must do worse on average over the remaining problems. [. . .] Thus, comparisons reporting the performance of a particular algorithm with a particular parameter setting on a few sample problems are of limited utility" [15]. For this reason, a wide range of ML regression algorithms has been implemented in this study: multiple linear regression (MLR), k-nearest neighbors (KNN), classification and regression tree

(CART), support vector regression (SVR), four ensemble methods (random forest, RF; gradient boosting, GB; AdaBoost, AB; extreme gradient boosting, XGB) and artificial neural networks (ANNs, in this case, multi-layer perceptron, MLP). A brief description of these algorithms is as follows:

- In MLR, the relationship between the predictors and the response variable is fitted through a multilinear equation to the observed data. MLR is considered as a baseline algorithm for regression, i.e., a simple model with a reasonable chance of providing decent results. Baseline models are easy to deploy and provide a benchmark to evaluate the performance of more complex models. Many early approaches to embrittlement modeling for RPV steels adopted a "chemistry factor" that was a weighted sum of the contributions of different chemical elements to embrittlement [16]. The data fittings needed to establish these chemistry factors were essentially multi-linear regressions.
- In KNN, regression is carried out for a new observation by averaging the target variable of the 'K' closest observations (the neighbors) with weights that can be uniform or proportional to the inverse of the distance from the query point. KNN is an example of an instance-based algorithm that depends on the memorization of the dataset; then, predictions are obtained by looking into these memorized examples. The distance between instances is based on features (e.g., Cu, Ni, temperature) important to the target variable (TTS) and is quantified through the Minkowski metric, which depends on the power parameter, 'p'. When $p = 1$, this is equivalent to using the Manhattan distance and for $p = 2$, the Euclidean distance.
- CARTs were introduced in 1984 by Breiman et al. [17]. These "trees" may be thought of as flowcharts; they have a branching structure in which a series of decisions is used to make a prediction that either classifies the data (outputting a categorical value) or regresses the data (outputting a continuous value). The CART splits the dataset to form a tree structure; the branching decisions (called decision nodes) are guided by the homogeneity of data. The resultant tree has "leaf nodes" at the end of the branches. Ideally, each leaf represents a more-or-less uniform response of the target variable, be it categorical or continuous. The Gini index and the entropy are the most common scores to measure the homogeneity of the data and to decide which feature should be selected for the next split. Building a decision tree requires finding the attribute that returns the highest information gain (which is defined as the entropy of the parent node minus the entropy of the child nodes after the dataset has been split on that attribute) or the highest reduction in the Gini index. The main advantages of decision trees are that the interpretation of results is straightforward and that they implicitly perform a feature selection since the earliest (or top) nodes of the tree are the most important variables within the dataset. The main limitation of a CART is that when a decision tree grows and becomes very complex, it usually displays a high variance and a low bias, which are evidence of overfitting. This makes it difficult for the model to generalize and to incorporate new data.
- The support vector machine (SVM) algorithm was originally designed as a classifier [18] but may also be used for regression, SVR, and feature selection [19]. In classification, SVM determines the optimal separating hyperplane between linearly separable classes maximizing the margin, which is defined as the distance between the hyperplane and the closest points on both sides (the support vectors). For non-perfectly separable classes, SVM must be modified to allow some points to be misclassified, which is achieved by introducing a "soft margin" [20]. Datasets that are highly nonlinear may in some cases be (linearly) separated after being (nonlinearly) mapped into a higher dimensional space [21]. This mapping gives rise to the kernel, which can be chosen by the user among different options such as linear, sigmoid, Gaussian or polynomial. The appropriate kernel function is selected by trial and error on the test set. In this case, SVM is referred to as kernelized SVM.
- Ensemble learning is a paradigm that focuses on training a large number of low-accuracy models, which are called "weak learners," and combining their predictions

to obtain a high accuracy meta-model. Decision trees (as just described under CARTs) are the most widely used weak learners. The idea behind ensemble learning is that if the trees are not identical and they predict the target variable with an accuracy that is slightly better than random guessing, a prediction based on some sort of weighted voting of a large number of such trees will improve accuracy. Ensemble methods are classified into bagging-based and boosting-based, which are designed to reduce variance and bias, respectively.

- Bagging (which stands for bootstrap aggregation) is the application of the bootstrap procedure (i.e., random sampling with replacement) to a high-variance ML model (e.g., a regression tree). Many models are created, and every model is trained in parallel. Each of the models is trained on a subset of the whole dataset composed of a number of observations randomly selected with replacement using a subset of features. The predicted value produced by bagging is simply the average of the predictions from all the models. The most widely used bagging-based ML algorithm is RF [22], which uses shallow classification trees as the weak learners. The most important hyperparameters to tune in a RF are the number of trees and the size of the random subset of the features to consider at each split. By using multiple samples of the original dataset, the variance of the ensemble RF model is reduced, as is the overfitting (see Section 2.3.4).
- Boosting consists of using the original training data and iteratively creating multiple models by using a weak learner, usually a regression tree. Each new model tries to fix the errors made by previous models. Unlike bagging, which aims at reducing variance, boosting is mainly focused on reducing bias. In adaptative boosting, AB, and in GB, the ensemble model is defined as a weighted sum of weak learners. The weights are placed more heavily on the data that were poorly predicted by the initial weak learners, thereby gradually improving the accuracy of the overall model. XGB is an algorithm that uses a GB framework developed in 2016 by Chen and Guestrin [23]. Among its advantages, it provides a good combination of performance and processing time through systems optimization and algorithmic enhancements (such as parallelized implementation).

- ANNs are used for data classification, regression and pattern recognition. A basic ANN contains a large number of neurons arranged in layers. An MLP begins with an input layer, which contains one or more hidden layers that are trained to make decisions, and an output layer. The nodes of consecutive layers are connected, and these connections have weights associated with them. During training, weights are initially assigned randomly. Known data are then fed forward through the network from the input nodes, through the hidden nodes (if any), and to the output nodes. The output of every neuron is obtained by applying an activation function to the linear combination of inputs (weights) to the neuron; sigmoid, tanh and rectified linear unit (ReLu) are the most widely used activation functions. MLPs are trained to produce better answers through backpropagation. During backpropagation, the network is provided with feedback concerning outputs that have been incorrectly predicted. The network then changes the weights associated with the nodes in the hidden layers to produce a more accurate output. Gradient descent, Newton, conjugate gradient and Levenberg–Marquardt are different algorithms used to train an ANN.

2.3.4. Evaluation of Machine Learning Algorithms

Before undertaking any preprocessing [11,12], 25% of the observations were randomly extracted to form a test dataset that was used to provide an unbiased evaluation of the performance of the models. Models were trained, and the hyperparameters were refined using the remaining 75%, i.e., the train dataset. This train/test separation was conducted by stratifying the two categorical variables (REACTOR_TYPE and PRODUCT_ID) to

ensure the representativeness of both the train and test datasets to the reactor types and product forms represented by the PLOTTER BASELINE data. Threefold cross-validation, conducted with GridSearchCV, was used for hyperparameter tuning; these concepts are briefly described in the next paragraph.

The hyperparameters of an ML model are those parameters that are determined prior to training, either specified by the practitioner or using some heuristics. For example, in a parametric model such as polynomial regression, the degree of the polynomial has to be set somehow before training; then, after training, the coefficients of the polynomial (which represent the parameters of the model) are obtained. The most commonly used method is grid search, which consists of providing a list of values for each hyperparameter of interest and trying all possible combinations among them to finally select the optimal one [11,12] (see Figure 4). To make this decision, an option would be to choose the combination providing best scores on the 25% test set described in the previous paragraph; however, in that case, the test set should not be used to assess how good the model is because it is potentially contaminated with information used for training (in this case, for training to set the hyperparameters). An alternative approach to resolve this problem is to split the data into three sets: the 75% training set is itself divided into two parts: one part is used to build the model while another part (the validation set) is used to tune the hyperparameters of the model. The 25% test set is used to obtain an unbiased estimation of the actual performance of the model on fresh data. The K-fold cross validation process achieves Step A. Here, the 75% training set test is randomly split into K distinct subsets called folds. At each iteration, a different fold is chosen as the validation set, and the model is trained on the other K-1 folds; this process is schematically depicted in Figure 4.

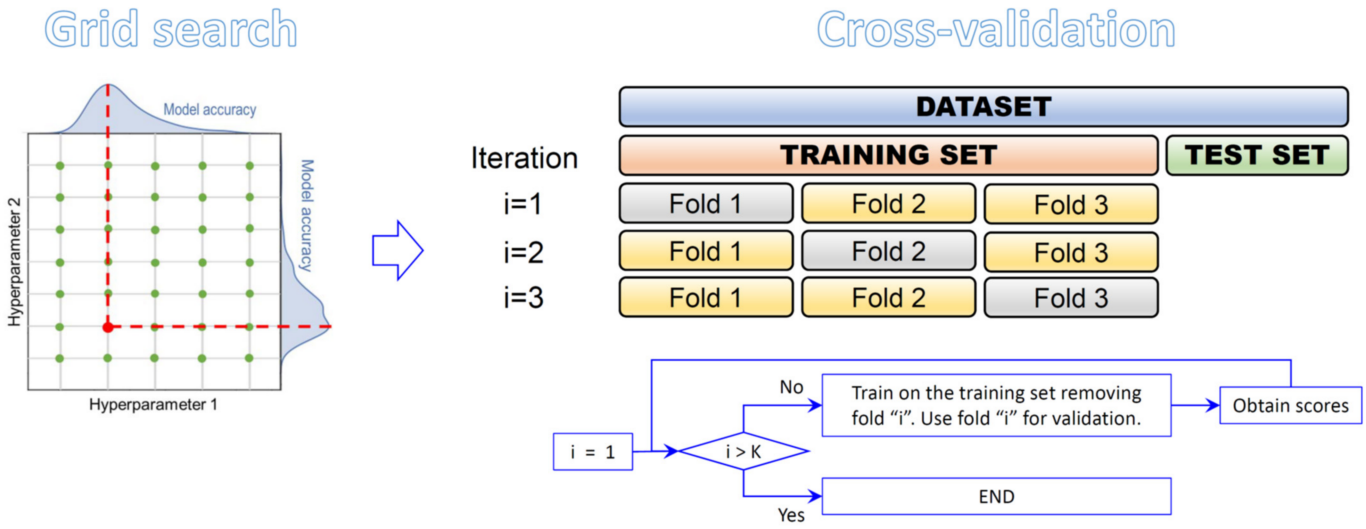


Figure 4. Schematic description of the procedure followed in this research for hyperparameter tuning. Grid search was used to obtain the combinations of the values of the hyperparameters. Then, a threefold cross validation method was implemented for hyperparameter tuning.

The goal of supervised learning is to build a model on the training data to make accurate predictions on new, unseen data (provided they have the same characteristics as the training set). When a model makes accurate predictions on unseen data, it is said that it generalizes from the train set to the test set. The test/train split evaluates the ML model accuracy by assessing the propensity of each ML model for both overfitting and underfitting [11,12].

If a model overfits the data, it performing well on the training data but less accurately on the test set. Complex models are likely to detect patterns in the noise itself. Obviously, these models will not generalize to new instances. Underfitting is the opposite of overfitting: it occurs when the model is too simple to represent the underlying structure of the data.

The following three scores were used to measure the quality of the regression ML model, namely, the coefficient of determination, R^2, the root mean square error, RMSE, and the mean absolute error, MAE.

3. Results

3.1. Selection of the Optimum Algorithm

In the first stage, the nine algorithms implemented were trained using the default options for their hyperparameters i.e., without conducting any tuning strategy (these default parameters can be consulted in the web page of Scikit-Learn). The results are shown in Table 1; the default hyperparameters are collected in Table 2. As can be seen, systematically the scores for the train set are better than for the test set. This is particularly so for those algorithms that are more prone to overfitting, such as the CART or the RF. As can be observed, the various metrics all convey the same trends. In absolute terms, best results in the test set are provided by the GB, XGB and RF algorithms, respectively. Figure 5 shows two scatterplots where the relative overfitting of the measures of scatter, RMSE and MAE, respectively, is represented against the relative overfitting in terms of the coefficient of determination. This representation shows that, in relative terms, CART, RF and KNN are the algorithms that display a larger overfitting. Taking these arguments into consideration, the GB algorithm was selected for hyperparameter tuning.

Table 1. Scores (R^2, RMSE and MAE) obtained in the train and test datasets for the nine algorithms implemented without tuning. MLR: multiple linear regression. KNN: k-nearest neighbors. CART: classification and regression tree. SVR: support vector regression. RF: random forest. AB: AdaBoost. GB: gradient boosting. XGB: extreme gradient boosting. MLP: multi-layer perceptron.

	Train Dataset			Test Dataset		
Regressor	**R^2**	**RMSE (°C)**	**MAE (°C)**	**R^2**	**RMSE (°C)**	**MAE (°C)**
MLR	0.736	18.87	18.87	0.725	20.91	15.34
KNN	0.881	12.66	12.66	0.810	17.39	12.51
CART	0.995	2.59	2.58	0.830	16.44	12.44
SVR	0.579	23.85	23.85	0.597	25.32	16.45
RF	0.972	6.10	6.10	0.872	14.26	10.34
AB	0.810	16.01	16.01	0.779	18.77	14.50
GB	0.927	9.94	9.94	0.896	12.87	9.81
XGB	0.924	10.10	10.10	0.888	13.36	10.06
MLP	0.874	13.03	13.03	0.863	14.76	10.76

Table 2. Hyperparameters used in the train and test datasets for the nine algorithms implemented without tuning.

Regressor	**Hyperparameters**
MLR	N/A
KNN	n_neighbors = 5, weights = 'uniform', algorithm = 'auto', leaf_size = 30, p = 2, metric = 'minkowski'
CART	criterion = 'squared_error', splitter = 'best', min_samples_split = 2, min_samples_leaf = 1
SVR	kernel = 'rbf', degree = 3, gamma = 'scale', tol = 0.001, C = 1.0, epsilon = 0.1, shrinking = True, cache_size = 200, verbose = False, max_iter = −1

Table 2. *Cont.*

Regressor	Hyperparameters
RF	n_estimators = 100, criterion = 'squared_error', min_samples_split = 2, min_samples_leaf = 1, max_features = 'auto', bootstrap = True, oob_score = False, verbose = 0, warm_start = False
AB	n_estimators = 50, learning_rate = 1.0, loss = 'linear'
GB	oss = 'squared_error', learning_rate = 0.1, n_estimators = 100, subsample = 1.0, criterion = 'friedman_mse', min_samples_split = 2, min_samples_leaf = 1, max_depth = 3, alpha = 0.9, warm_start = False, validation_fraction = 0.1, tol = 0.0001
XGB	'objective': 'reg:squarederror', 'importance_type': 'gain', 'n_estimators': 100
MLP	hidden_layer_sizes = (100), activation = 'relu', solver = 'adam', alpha = 0.0001, batch_size = 'auto', learning_rate = 'constant', learning_rate_init = 0.001, power_t = 0.5, max_iter = 200, shuffle = True, tol = 0.0001, verbose = False, momentum = 0.9, nesterovs_momentum = True, early_stopping = False, validation_fraction = 0.1, beta_1 = 0.9, beta_2 = 0.999, epsilon = 1×10^{-8}, n_iter_no_change = 10, max_fun = 15,000

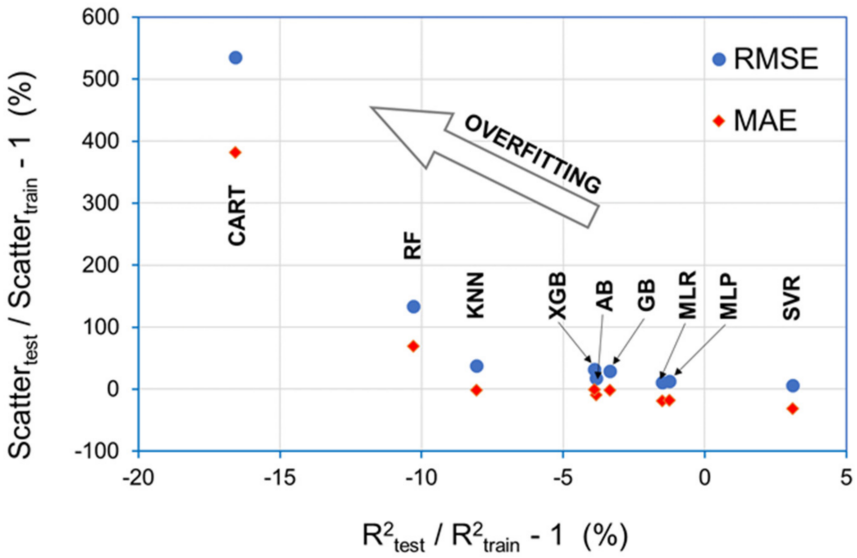


Figure 5. This scatterplot provides an estimation of the relative performance between the test and train datasets in terms of the coefficient of determination and the two scores employed to measure the scatter, the RMSE and MAE, respectively. The arrow superimposed on the figure shows the direction of increasing overfitting.

3.2. Regression with Gradient Boosting

Hyperparameters are parameters that are not directly learned within estimators. In this study, they were tuned for the GB algorithm using a grid search and cross-validation schemes to obtain the best values of R^2 (highest), RMSE and MAE (lowest) for the train dataset. The optimal hyperparameters are as follows: n_estimators = 500, max_depth = 4, learning_rate = 0.1, max_features = 8, min_samples_leaf = 5, min_samples_split = 3. As can be seen in Table 3, a moderate but acceptable amount of overfitting exists as evidenced by the inferior values of R^2, RMSE, and MAE for the test dataset, especially considering that all attempts to reduce the complexity of the model have also penalized the accuracy in the test dataset. For the sake of completeness, the normalized root mean squared error (NRMSE), which is defined as the RMSE divided by the standard deviation, has also been included in the table to provide a comparison between datasets with different scales.

Table 3. Scores (R^2, RMSE, NRMSE and MAE) obtained with the tuned GB model in the train, test and train + test datasets. The last column shows the scores obtained using the predictive model of ASTM E900-15.

	Train	Test	Train + Test	ASTM E900-15
R^2	0.977	0.914	0.963	0.875
RMSE (°C)	5.73	10.54	7.24	13.32
NRMSE	0.159	0.277	0.198	0.364
MAE (°C)	4.23	8.37	5.26	10.12

The scatterplots in Figure 6 represent the TTS predicted with the tuned GB model as a function of the experimental values for the train (a) and test (b) datasets. For comparison, the corresponding information obtained from the ASTM E900-15 model [4] has been represented in (c).

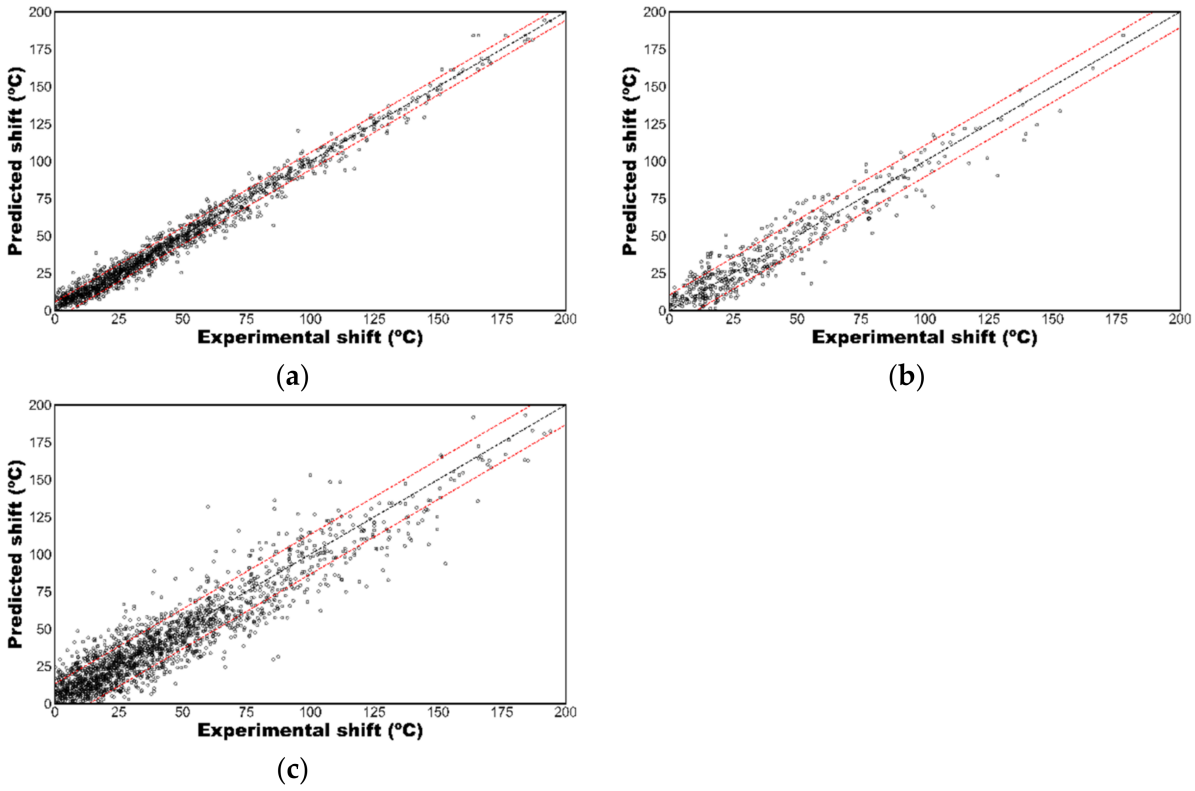


Figure 6. The scatterplots compare the experimental shift with the predicted values obtained from the optimized GB model in the train set (**a**) and test set (**b**). In (**c**), the predictions of the ASTM E900-15 [4] model are represented. All pictures have been represented with the same scale to facilitate the comparison. A 1:1 black dotted line as well as two red dotted lines vertically separated from the former by a distance equal to the RMSE have been included in the figure.

Figure 7 provides a comparison of the residuals obtained from the GB and the ASTM E900-15 models for the three sets previously mentioned. In all scenarios, the ML-based approach outperforms the analytical model of the ASTM E900-15 standard [4]. This is true even when the GB model is faced with fresh data, that is, data not used in the training process. Indeed, Figure 7b shows a substantial improvement for observations with higher residuals, those higher (in absolute value) to 25 °C (its presence is appreciable in the analytical model, but testimonial in the ML model).

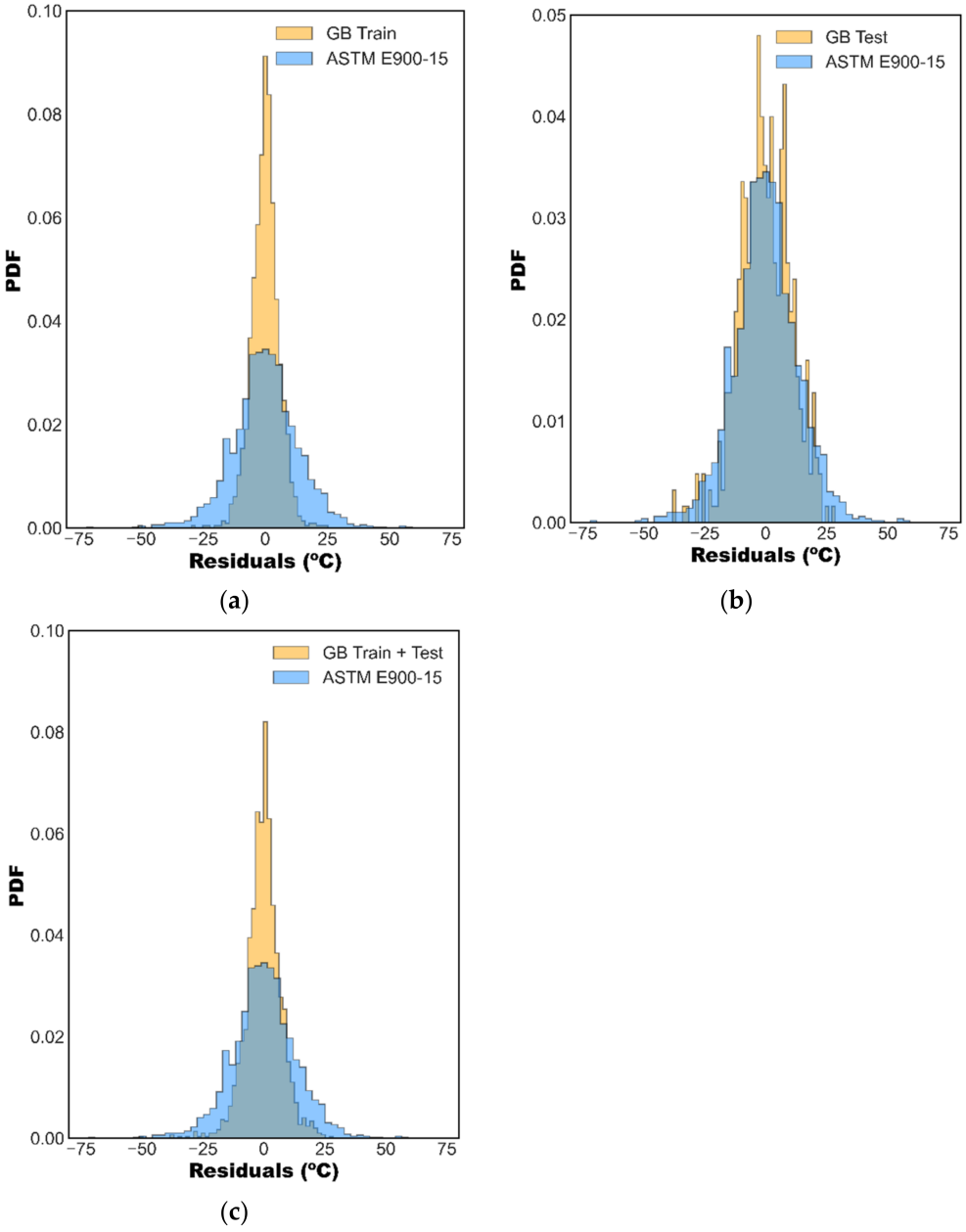
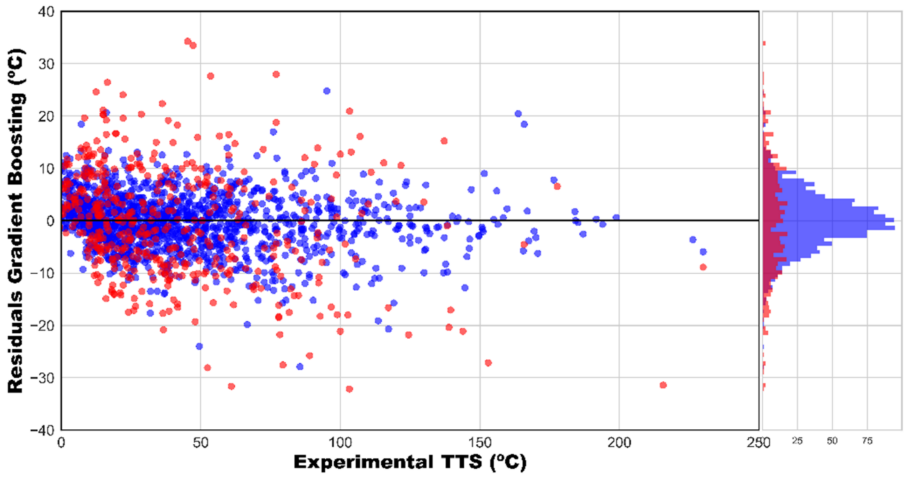
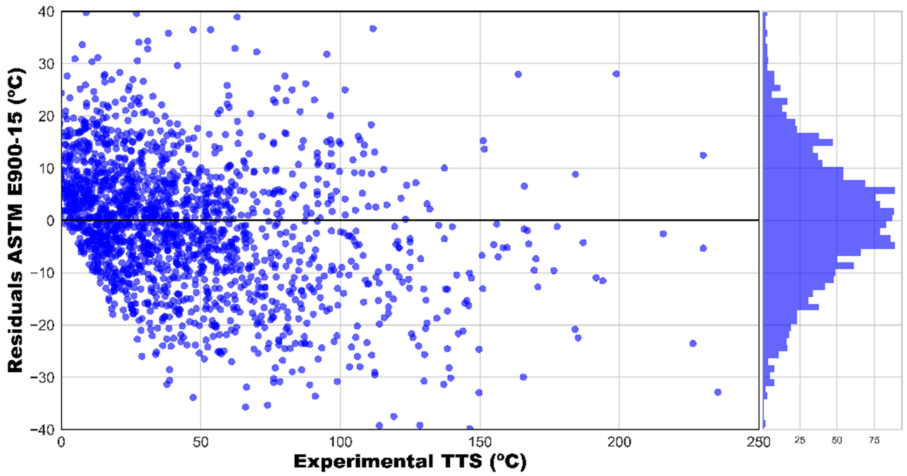


Figure 7. Comparison of residuals obtained from the GB algorithm and the ASTM E900-15 ETC. (**a**) Train dataset, (**b**) test dataset, (**c**) all data.

An examination of the distribution of residuals is always a good practice in descriptive statistics. The residuals plot provided by the GB algorithm for the train and test datasets represented in Figure 8a exhibits a random pattern, and no clear trend is appreciated. Particularly remarkable are the large residuals that can be seen for low values of the TTS, these being associated with the errors inherent in determining TTS from a limited number (generally 8–12) of Charpy tests. The mean value of the residuals is 0.3 °C, which indicates an unbiased estimation. For comparison, the residuals obtained using the ASTM E900-15 [4] ETC are represented in Figure 8b; in this case, the bias is −0.1 °C.



(**a**)



(**b**)

Figure 8. (**a**) Residuals plot obtained through the GB algorithm for the train and test datasets, (**b**) residuals plot provided by the ASTM E900-15 ETC.

To complete the graphical representations, in Figure 9a the experimental values and the predictions of the GB algorithm are compared again in the form of a scatterplot. In this case, the data have been represented as a color map based on the density of points, and contour lines have been superimposed. The relative scarcity of values with TTS > 60 °C (24.8% of the database) and the existence of observations having a large residual for small embrittlement are noticeable in both Figures 8 and 9. An equivalent scatterplot based on the predictions provided by the ETC of ASTM E900-15 [4] is shown in Figure 9b; as can be appreciated, the dispersion of data is significantly larger.

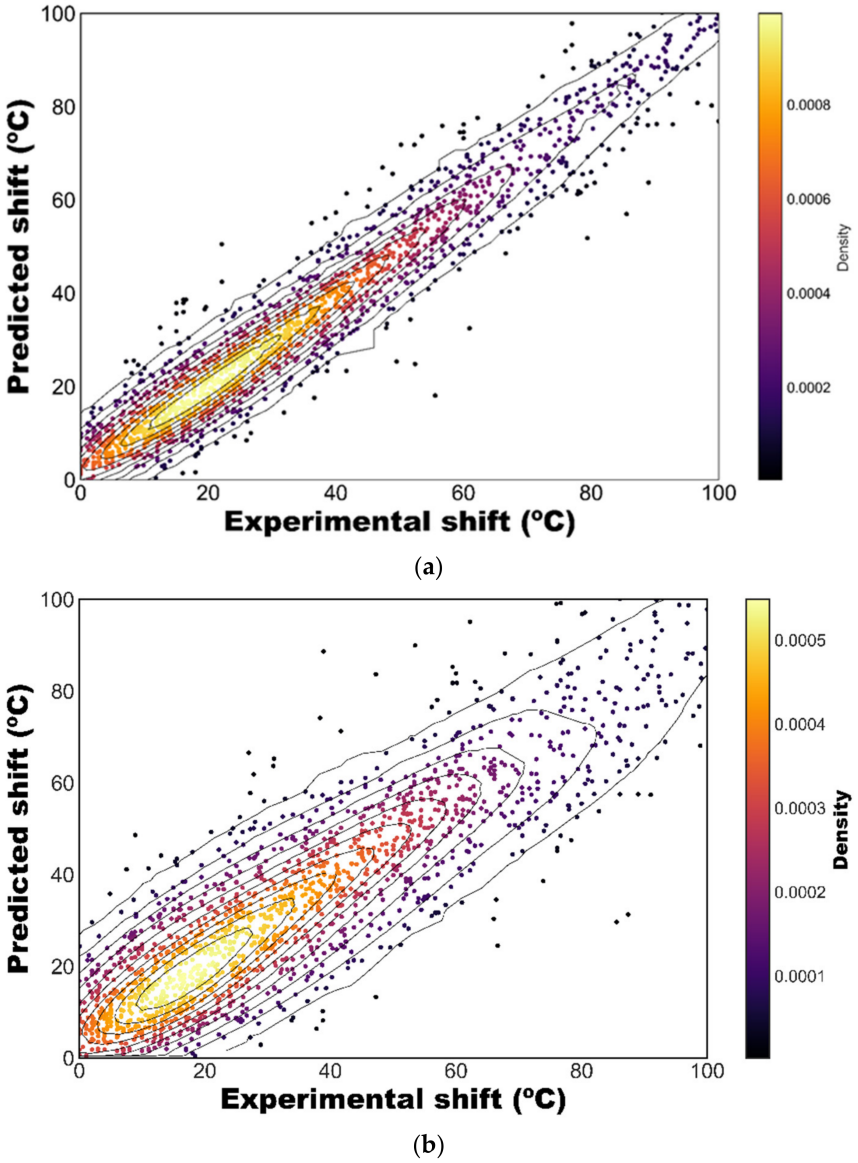
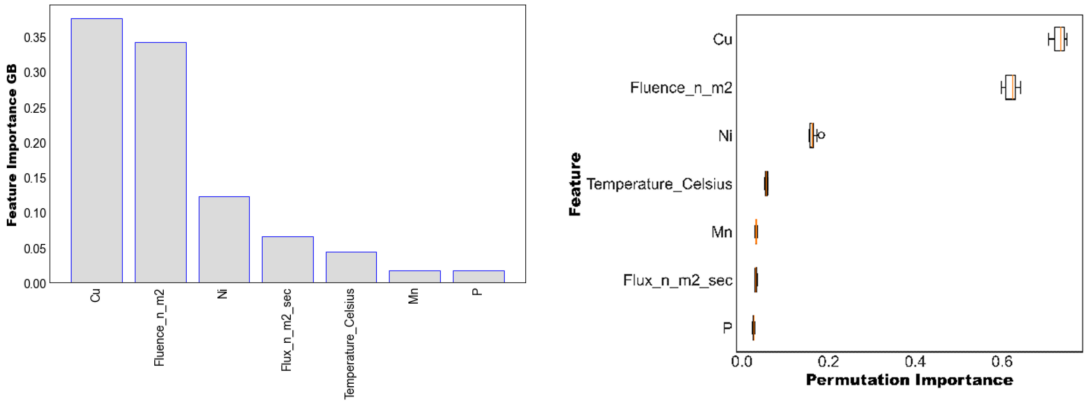


Figure 9. Scatterplot relating the experimental and predicted TTS. (**a**) Gradient Boosting, (**b**) ASTM E900-15 [4].

3.3. Assessment of Importance of the Features

Tree-based models such as GB provide metrics that can be used to evaluation the relative importance of the different features (e.g., copper, nickel, fluence) of the model. These metrics include a mean decrease in impurity, which is the splitting criterion [13] and a permutation feature importance, which is defined to be the decrease in a model score when a single feature value is randomly shuffled [13]. This procedure breaks the relationship between the feature and the target variable TTS, thus the drop in the model score is indicative of how much the model depends on the feature. This technique benefits from being model agnostic (it can be used with any model, not only CARTs or ensembles of trees) and can be calculated many times with different permutations of the feature [13].

Figure 10 compares the results obtained using the (a) impurity-based and (b) permutation-based feature importance. Even though the numeric values differ (because they are measuring different things), both procedures suggest that copper, fluence, and temperature are the most important features in estimating TTS, while features such as phosphorus and manganese play minor roles.

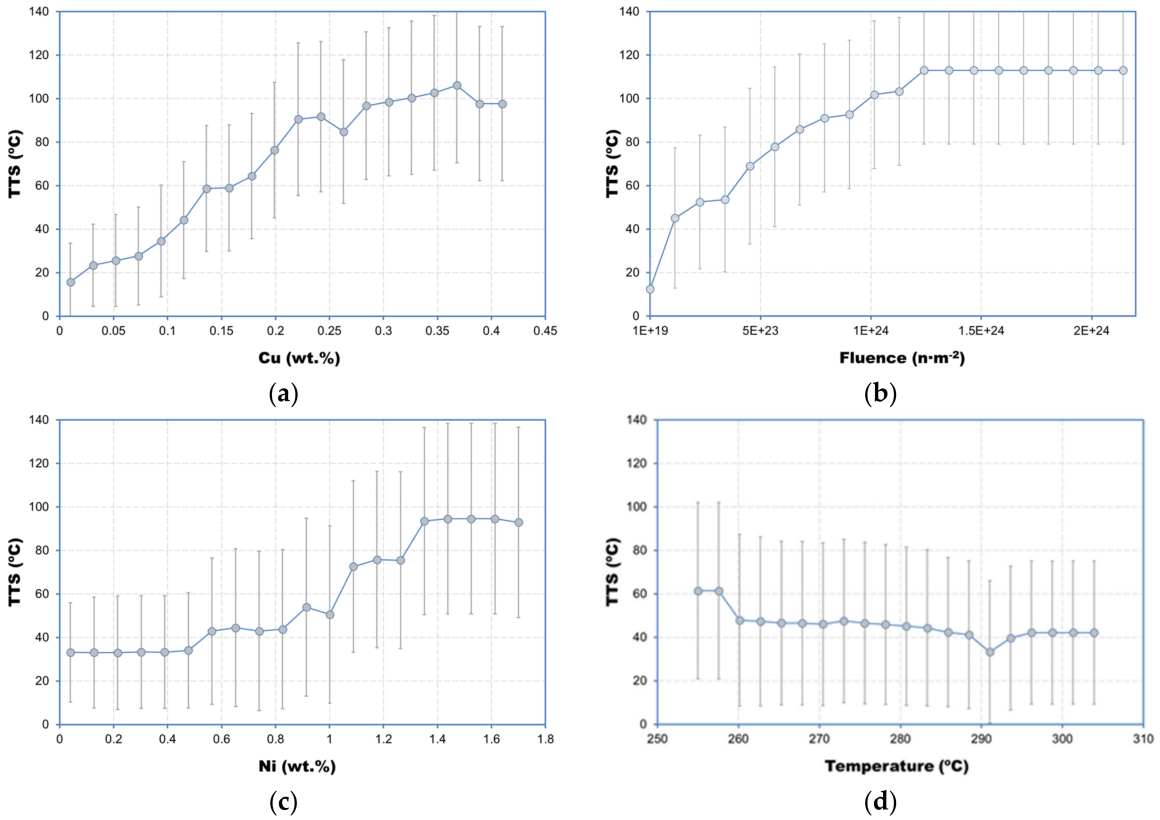


(**a**) (**b**)

Figure 10. (**a**) Impurity-based and (**b**) permutation-based feature importance obtained from the optimized GB model.

3.4. Individual Conditional Expectation Plots and Interaction Plots

An individual conditional expectation (ICE) plot shows the dependence between the target variable (the TTS in this case) and an input feature of interest, as predicted by the ML model selected (GB in this case). An ICE plot usually includes one line per sample, but here, to facilitate the examination of the figure, the distribution of samples around the central trend was represented using error bars (with a length equal to ±1 standard deviation) [13]. Figures are sorted in Figure 11 based on their permutation importance (see Figure 10b). As was seen in the feature importance plots (Figure 10), some features (copper, fluence, nickel, temperature) exert a much greater influence on the target (TTS) than do others (manganese, flux, phosphorus).



(**a**) (**b**)

(**c**) (**d**)

Figure 11. *Cont.*

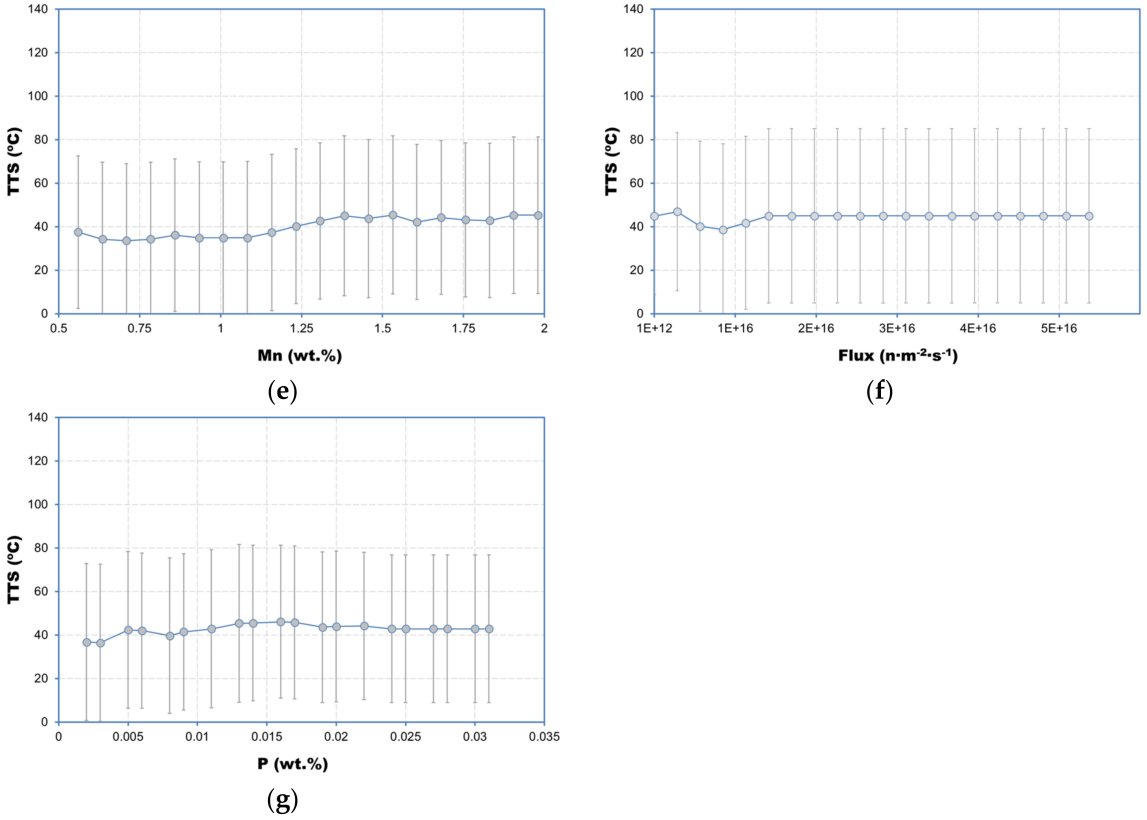


Figure 11. ICE plots showing the impact of each of the features of interest on the TTS. (**a**) Cu, (**b**) Fluence, (**c**) Ni, (**d**) Temperature, (**e**) Mn, (**f**) Flux and (**g**) P.

ICE plots describe the average contribution of each feature to the target response but ignore the possible interaction between features. Two-dimensional (2D) interaction plots [24] enable visualization of interactions among couples of attributes; the interaction between two features quantifies the change in the prediction that occurs by varying the features after considering the individual feature effects. Figure 12 collects the 2D interaction plots, represented through color maps, between the couples existing between the four most important features: Cu, fluence, Ni and T (the rest of combinations have not been represented because in these cases, one of the four previously mentioned features clearly overwhelms the other one). Each color in Figure 12 represents the same TTS ("iso-TTS" regions). Some basic ideas may help to understand these figures. Suppose that features 'X' and 'Y' are being represented in the horizontal and vertical axes, respectively. Vertical iso-TTS lines mean that feature Y does not exert any influence on the TTS; conversely, horizontal iso-TTS lines are associated with a negligible impact of the X feature.

For the sake of clarity and completeness, Figure 13 shows a magnified version of the 2D interaction plot between the two most relevant features, Cu and fluence. Several regions can be distinguished in the map. Thus, for fluence $< 10^{23}$ n $\times$ m^{-2}, region 1, embrittlement is basically dominated by the fluence, regardless of the copper content (iso-TTS lines are approximately horizontal in this zone). For fluence $> 10^{23}$ n $\times$ m^{-2} and Cu $< 0.12\%$, region 2, there is a clear interaction between these features. Finally, the behavior in region 3 is significantly noisy, although the influence of the fluence on the TTS seems to prevail. This analysis is but an example of the type of information that can be obtained from a 2D interaction plot.

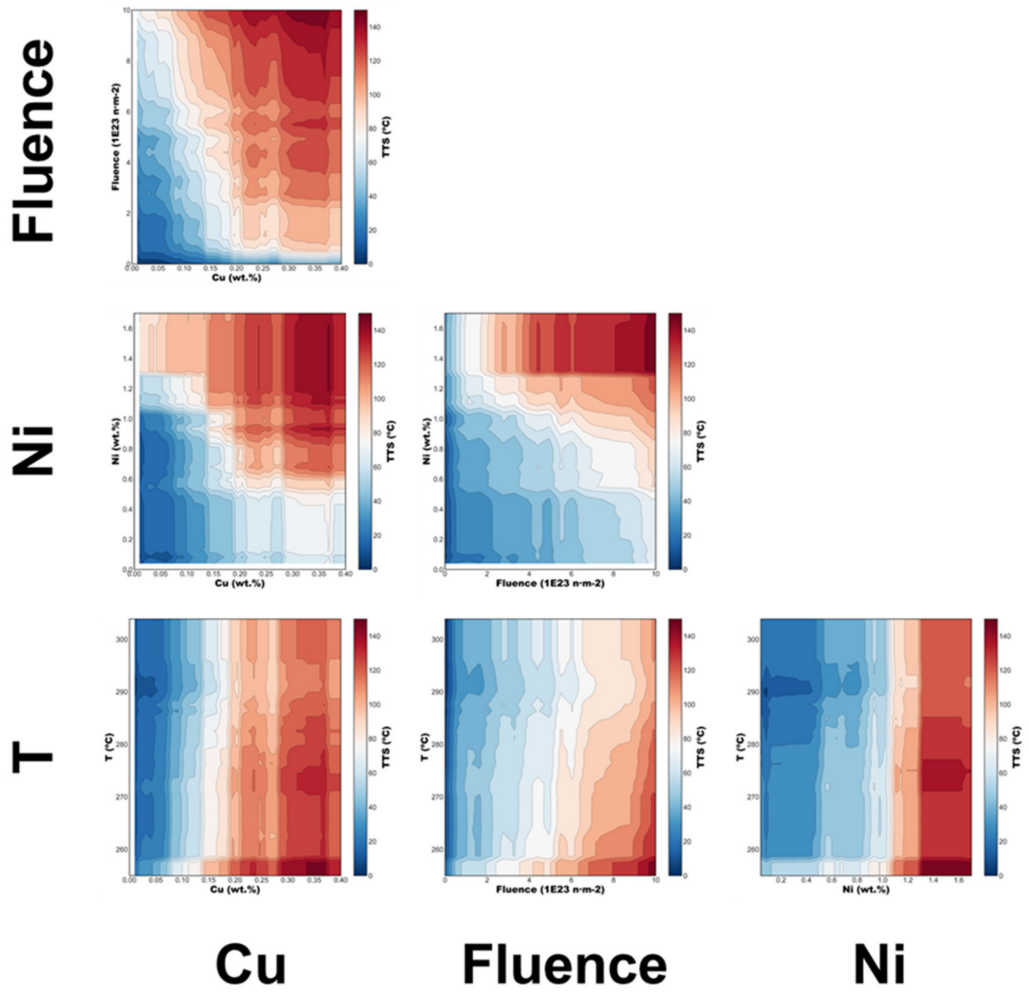


Figure 12. 2D Interaction plots between the four most relevant features: Cu, fluence, Ni and T.

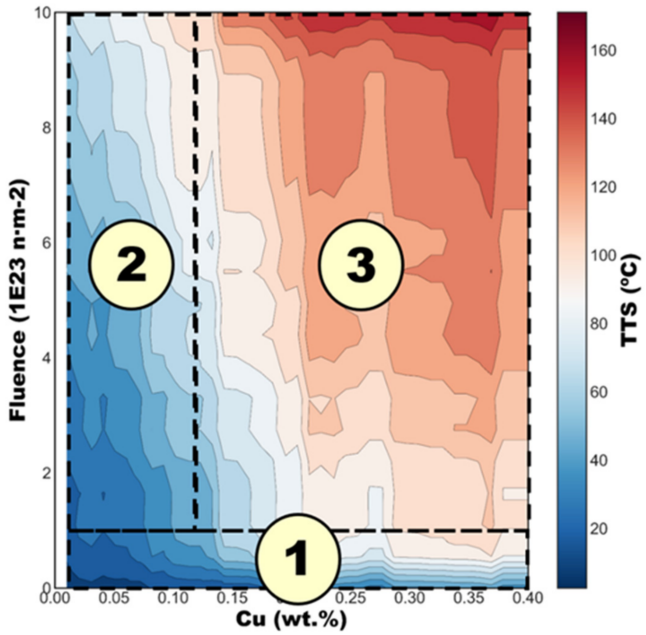


Figure 13. D Interaction plots between the two most relevant features, Cu and Fluence.

4. Discussion

Invoking the "No Free Lunch" theorem [14,15], a number of ML algorithms were trained for this study. The scores (R^2, RMSE and MAE) obtained in the test dataset demonstrate that, in general, good results are obtained by means of ensemble-type algorithms (RF, XGB and, specifically, GB) as well as with the MLP. Without questioning the validity of the "No Free Lunch" theorem, experience shows that ensemble methods provide highly competitive results in shallow learning settings (for example, this can be verified in the popular ML competition website Kaggle, http://kaggle.com, last access on 12 January 2022). As stated by Chollet [25] regarding GB, "the use of the GB technique results in models that strictly outperform RF most of the time, while having similar properties. It may be one of the best, if not the best, algorithm for dealing with nonperceptual data today."

The scores (R^2, RMSE, MAE) obtained through the optimized GB algorithm were (0.977, 5.73, 4.23) in the train set, (0.914, 10.54, 8.37) in the test set and (0.963, 7.24, 5.26) in the combined set. Therefore, strictly speaking, the model displays a *whiff* of overfitting. To correct for overfitting, a series of attempts were made to improve the generalizability of the model (i.e., reducing the differences observed between the scores in the train and test sets) by decreasing its complexity. All of them were unsuccessful and reduced the predictive capacity of the model in the test set; for this reason, these alternatives were rejected. To evaluate this model, it is important to consider the intrinsic uncertainty in a TTS measurement. The uncertainty in determining T_{41J} from a set of 8–12 Charpy data is between 5–6 °C [26], which makes the experimental uncertainty on the TTS approximately between $(5^2 + 5^2)^{0.5} \approx 7.1$ °C and $(6^2 + 6^2)^{0.5} \approx 8.5$ °C, which is an absolute lower limit for any (not overfitted) ETC. By comparison, the RMSD for the test dataset was 10.5 °C. This value exceeds 7–8.5° C, suggesting that overfitting is not a significant issue. The RMSD for the GB model developed herein was also 21% lower than for the analytical model of ASTM E900-15, suggesting good improvement in prediction accuracy.

Lee et al. [27] previously developed an ETC using ML methods. They trained several regression models using the ASTM PLOTTER database described in Section 2.1. They implemented the XGB, Cubist and SVM algorithms and used the RMSE to evaluate the performance of the models. Best results were obtained with XGB (which agrees well with the rationale exposed above emphasizing the general superiority of GB for shallow learning), with RMSE = 11.9 °C. This result outperforms the prediction ability of the ASTM E900-15 nonlinear regression model, where RMSE = 13.3 °C. Thus, Lee's results and those presented herein are comparable

It is hard to overstate the importance of properly splitting the available data into three sets: training, validation, and test. The model is trained on the training data and its hyperparameters are tuned on the validation set. Finally, the actual unbiased quality of the model is measured on the test data. This procedure avoids information leaks because during hyperparameter tuning, significant amounts of information may be leaked from the validation set into the model, producing a regressor/classifier that performs artificially well on the validation data but whose performance on completely new data is unknown [11,12,25].

According to the values of impurity-based and permutation-based feature importance, four of the features exert a clear influence on the TTS: Cu, fluence, Ni and temperature. The ICE plots depicted in Figure 11 can be intuitively interpreted as the expected target TTS as a function of the input features of interest [13]. Therefore, they provide valuable information about the patterns associated with each variable. The ICE plot of Cu in Figure 11a shows a marked influence on the embrittlement of the material. Other things being equal, the average difference between steels with low and high Cu content can be greater than 80 °C of TTS. Furthermore, it is possible to glimpse three regions on the ICE curve. A first region for Cu < 0.075% with a moderate slope, a second approximately linear embrittlement region for 0.075% < Cu < 0.275%, and a final saturation region when Cu > 0.275%. This result is very consistent with the regression model of the ASTM E900-15 standard [4], which describes the influence of copper with a linear model for the region defined by the inequality 0.053%

< Cu < 0.28%, as well as with the Kirk model [28], which considers a linear response in the interval 0.07% < Cu < 0.30%. In both cases, the influence of Cu is considered null out of this range. The average influence of fluence, Figure 11b, can lead to differences in TTS greater than 100 °C between steels exposed to low or high levels of neutron irradiation. To facilitate the inspection of the figure, it has been represented on a logarithmic scale (see Figure 11b). As can be seen, the impact of fluence is markedly linear in this representation in the range between 3.38×10^{23} n/m^2 and 1.24×10^{24} n/m^2 (the logarithmic fitting conducted in this interval provides an $R^2 = 0.9859$). When fluence is out of this range, its influence seems to be reduced. Specifically, the ICE plot shows a plateau beyond 1.24×10^{24} n/m^2: this result must be interpreted carefully because, as previously noted by Lee et al. [27], extrapolation for tree-based ML models (such as GB) provides a constant value. Again, it is worth noting that both the analytical models proposed by the ASTM E900 [4] and by Kirk [28] describe the impact of fluence on the TTS through a logarithmic expression. In turn, the influence of Ni, as represented in Figure 11c, can also be described by means of a linear model restricted to an interval, in this case, approximately 0.5% < Ni < 1.3%. Outside of this range, its influence seems negligible. Finally, as suggested by Figure 11d, all else being equal, on average a higher temperature produces a lower TTS: this effect is moderate (on average, −0.32 °C/°C with a 95% confidence interval of (−0.46, −0.19) °C/°C)), but statistically very significant (the *p*-value for the slope is $p = 7.10 \times 10^{-05}$). The interpretation of the rest of the variables, Mn, flux and P, may not be so evident in a visual inspection. To assess their impact on the TTS, a regression linear fitting (on a semi-logarithm plot) of the mean ICE (see Figure 14) was carried out for these attributes. The p-values of the corresponding slopes were, respectively, p(Mn) = 4.084×10^{-07}, p(Flux) = 5.081×10^{-07} and p(P) = 2.53×10^{-02}. Therefore, in all cases, they were statistically significant at the 0.05 significance level. However, in all cases too, the slopes were small, producing a TTS change of 9 °C per percentage increase in Mn, 12 °C per 1.0×10^{16} n·m^{-2}·s^{-1} of flux and 14 °C per percentage increase in P.

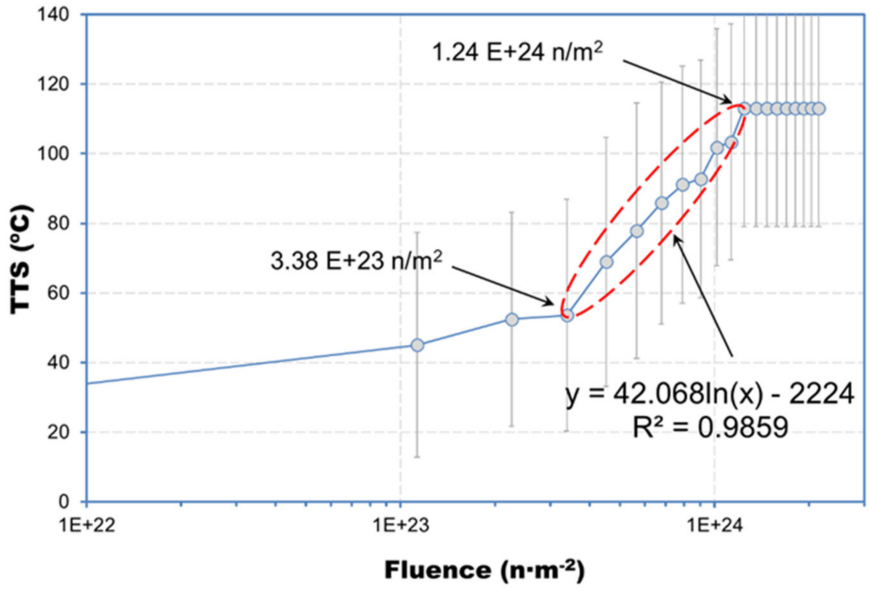


Figure 14. The graph shows the average influence of the fluence on the TTS. A linear relation with $R^2 = 0.9859$ can be observed for fluences between 3.38×10^{23} and 1.24×10^{24} n m^{-2}.

Figure 12 shows noticeable interactions between the following couples of attributes: Cu-fluence, Cu-Ni and fluence-Ni. In the case of Cu and fluence, it is observed that, except for low values of the fluence (less than ~0.7×10^{23} n·m^{-2}), Cu exerts a minor influence on the TTS and the embrittlement is dominated by the level of irradiation. For fluences above that limit, two regions can be distinguished: when Cu is approximately lower than 0.10–0.15%, it plays an important role in the embrittlement of the steel. On the other hand,

for Cu > 0.15%, it is the fluence that mostly explains the TTS. With regards to the interaction between Cu and Ni, the influence of Ni is only significant for Ni > 0.5%; specifically, in the range 0.5 < Ni < 1.1%, the value of the TTS is influenced by the content of Ni when Cu > 0.20%, approximately. A similar pattern can be appreciated for the interaction between fluence and Ni. The latter plays a role when 1.1 < Ni < 1.3%. Finally, no interaction seems to be observed between temperature and any of Cu, fluence or Ni.

To finish this discussion, it is necessary to note the possible limitations of our study. First, as mentioned above, the dataset we used comprises the 1878 observations of the PLOTTER database obtained from the surveillance programs of nuclear power reactors. Therefore, our model includes only a limited quantity of observations with very high values for fluence or flux, such as those that can be obtained in experimental reactors. Regarding the representativeness of PLOTTER to make predictions for long-term operation, this dataset includes fluences up to ~1×10^{24} n·m^{-2}, with an acceptable population of observations up to ~5×10^{23} n·m^{-2}. As reported by ASTM, see [29], the fluence achieved at the end of a particular operational duration depends on several factors, including, among others, reactor design, fuel loading and capacity factor. In [29], estimates of 80-year fluences for reactors now operating in the USA are given for BWRs that range from ~2×10^{21} to 2×10^{23} n/m^2 (median value ~3×10^{22} n/m^2), and for PWRs that range from ~2×10^{23} to 2×10^{24} n/m^2 (median value ~7×10^{23} n/m^2). Thus, while some reactors will experience fluences at the upper end of the range where surveillance data now exist, many will not. Likewise, [29] reports note that some new reactor designs will experience high fluences (e.g., the Westinghouse AP-1000 has a 60-year design fluence of ~9×10^{23} n/m^2). These figures guarantee the representativeness of the range of fluence values in the PLOTTER database and, at the same time, demonstrate the need to incorporate new data with higher fluence to guarantee an adequate assessment of structural integrity under very long-term operation. Undoubtedly, information coming from experimental reactors could be of great value to obtain models based on machine learning algorithms to develop embrittlement trend curves that provide more robust predictions for high fluence power plants during life extension. In a complementary way, this underlines the importance of increasing the population of observations available in PLOTTER because, as pointed out earlier, roughly 75% of the data correspond to TTS values under 60 °C.

From a practical perspective, it is worth considering the possibility of developing codes and standards or regulatory guidance based on machine learning algorithms. The information presented herein, as well as the similar study by Lee [22], suggests that modern machine learning tools can achieve improved predictive accuracy relative to the analytical ETC models traditionally used. Specifically, our study showed a 20% improvement in TTS prediction accuracy relative to the ASTM E900-15 analytical model. One approach would be to provide the complete code in Python; however, this may be impractical and restrict potential users to those skilled in this programming languages as well as in machine learning methods. Alternatively, a friendly application accessible for any user could be developed. This approach has been used by one of the authors of this paper in a recent contribution [30] where an application was created on the Microsoft.Net platform.

It should, however, be noted that from a theoretical perspective, predicting the TTS of new observations is not a straightforward task, whether using ML or more conventional analytical techniques. Kirk and Todeschini [31] have pointed out that the distribution of the data available in PLOTTER is sparse in some regions, which may hinder the ability to make predictions for specific plants. Different solutions can be implemented to avoid blind decision-making. For example, Kirk et al. [32] have analyzed the possibility of developing a local fitting using the KNN algorithm. In addition, there are specific methods developed to decide whether a new observation belongs to the same distribution as existing observations (it is an inlier), or should be considered as different (it is an outlier). In the latter case, the application of the developed model may be questionable and/or counterproductive.

5. Conclusions

The ASTM PLOTTER database that was used to inform the ASTM E900-15 TTS was employed to train and validate a number of machine learning regression models (multilinear, k-nearest neighbors, decision trees, support vector machines, random forest, AdaBoost, gradient boosting, XGB, and multi-layer perceptron). The main conclusions are:

- Best results in the test set were provided by the GB algorithm, which provided a value of R^2 = 0.91 and a root mean squared error ≈10.5 °C for the test dataset. These results outperformed the prediction ability of existing trend curves, including ASTM E900-15, reducing the prediction uncertainty by ≈20%.
- The impurity-based and the permutation-based feature importances obtained from the optimized GB model suggest that copper, fluence, and temperature are the most important features in estimating TTS, while features such as phosphorus and manganese play minor roles, in accordance with existing atomistic, physical and empirical models on neutron embrittlement.
- Noticeable interactions between Cu-fluence, Cu-Ni and fluence-Ni were observed
- The individual conditional expectation (ICE) plot for Cu yielded a classification of low Cu and high Cu consistent with the regression model of the ASTM E900-15.

Author Contributions: Conceptualization, D.F., M.K. and M.S.; methodology, D.F., M.K., M.S. and J.A.S.-A.; resources, M.S.; data curation, M.K. and J.A.S.-A.; writing—original draft preparation, D.F.; writing—review and editing, M.K., M.S. and J.A.S.-A.; funding acquisition, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work received partial financial support in the frame of the Euratom research and training programme 2019–2020 under grant agreement No 900018 (ENTENTE project).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ΔYS	Yield strength increase
AB	AdaBoost
ANN	Artificial neural networks
BWR	Boiling water reactor
CART	Classification and regression tree
ETC	Embrittlement trend curve
GB	Gradient boosting
ICE	Individual conditional expectation
KNN	K-nearest neighbors
LTO	Long-term operation
LWR	Light water reactor
MAE	Mean absolute error
ML	Machine learning
MLP	Multi-layer perceptron
MLR	Multiple linear regression
MTR	Material test reactor
NRMSE	Normalized root mean square error
PWR	Pressurized water reactor
r	Sample Pearson correlation coefficient
R^2	Coefficient of determination
ReLu	Rectified linear unit
RF	Random forest

RMSE	Root mean square error
RPV	Reactor pressure vessel
SVR	Support vector regression
TTS	Transition temperature shift
XGB	Extreme gradient boosting
VVER	Water-water energetic reactor
z	z-score

References

1. Eason, E.D.; Wright, J.E.; Odette, G.R. *Improved Embrittlement Correlations for Reactor Pressure Vessel Steels*; Division of Engineering Technology, Office of Nuclear Regulatory Research, US Nuclear Regulatory Commission: Washington, DC, USA, 1998.
2. *Regulatory Guide 1.99 (Revision 2): Radiation Embrittlement of Reactor Vessel Materials*; USNRC: Washington, DC, USA, 1998.
3. Eason, E.D.; Odette, G.R.; Nanstad, R.K.; Yamamoto, T. *A Physically Based Correlation of Irradiation-Induced Transition Temperature Shifts for RPV Steels*; U.S. Nuclear Regulatory Commission: Oak Ridge, TN, USA, 2007.
4. *ASTM E900-15e2, Standard Guide for Predicting Radiation-Induced Transition Temperature Shift in Reactor Vessel Materials*; ASTM International: West Conshohocken, PA, USA, 2015.
5. Hashimoto, Y.; Nomoto, A.; Kirk, M.; Nishida, K. Development of new embrittlement trend curve based on Japanese surveillance and atom probe tomography data. *J. Nucl. Mater.* **2021**, *553*, 153007. [CrossRef]
6. *The Fourth Paradigm: Data-Intensive Scientific Discovery*; Hey, T.; Tansley, S.; Tolle, K. (Eds.) Microsoft Research: Washington, DC, USA, 2009.
7. Unger, J.F.; Könke, C. Neural networks as material models within a multiscale approach. *Comput. Struct.* **2009**, *87*, 1177–1186. [CrossRef]
8. Zopf, C.; Kaliske, M. Numerical characterisation of uncured elastomers by a neural network based approach. *Comput. Struct.* **2017**, *182*, 504–525. [CrossRef]
9. Kalidindi, S.R.; Niezgoda, S.R.; Salem, A.A. Microstructure informatics using higher-order statistics and efficient data-mining protocols. *J. Miner.* **2011**, *63*, 34–41. [CrossRef]
10. Rajan, K. Materials informatics. *Mater. Today* **2005**, *8*, 38–45. [CrossRef]
11. Guido, S.; Müller, A. *Introduction to Machine Learning with Python: A Guide for Data Scientists*; O'Reilly Media: Newton, MA, USA, 2016; ISBN 978-1449369415.
12. Geron, A. *Hands-On Machine Learning with Scikit-Learn and TensorFlow*; O'Reilly Media, Inc.: Newton, MA, USA, 2017; ISBN 978-1491962299.
13. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M. Scikit-learn. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
14. Wolpert, D.H. The Supervised Learning No-Free-Lunch Theorems. In Proceedings of the 6th Online World Conference on Soft Computing in Industrial Applications, On the Internet (World Wide Web), 10–24 September 2001; Springer: Berlin/Heidelberg, Germany, 2001; pp. 1–20.
15. Wolpert, D.H.; Macready, W.G. No free lunch theorems. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82. [CrossRef]
16. *Regulatory Guide 1.99 (Revision 0): Effects of Residual Elements on Predicted Radiation Damage to Reactor Vessel Materials*; USNRC: Washington, DC, USA, 1975.
17. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and Regression Trees*; Chapman and Hall/CRC: London, UK, 1984; ISBN 978-0412048418.
18. Vapnik, V.; Chervonenkis, A. A note on one class of perceptrons. *Autom. Remote Control* **1964**, *25*, 61–68.
19. Cao, F.L.; Bai, H.B.; Yang, J.C.; Ren, G.Q. Analysis on Fatigue Damage of Metal Rubber Vibration Isolator. *Adv. Mater. Res.* **2012**, *490–495*, 162–165. [CrossRef]
20. Mohamed, A.E. Comparative Study of Supervised Machine Learning Techniques for Intrusion Detection. In Proceedings of the Fifth Annual Conference on Communication Networks and Services Research (CNSR'07), Frederlcton, NB, Canda, 14–17 May 2007; Volume 14, pp. 5–10. [CrossRef]
21. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992; pp. 144–152. [CrossRef]
22. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
23. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
24. Goldstein, A.; Kapelner, A.; Bleich, J.; Pitkin, E. Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation. *J. Comput. Graph. Stat.* **2015**, *24*, 44–65. [CrossRef]
25. Chollet, F. *Deep Learning with Python*; Manning Publications: New York, NY, USA, 2018; ISBN 978-1617294433.
26. Soneda, N.; Dohi, K.; Nomoto, A.; Nishida, K.; Ishino, S. Embrittlement Correlation Method for the Japanese Reactor Pressure Vessel Materials. In *Effects of Radiation on Nuclear Materials and the Nuclear Fuel Cycle: 24th Volume*; Busby, J., Hanson, B., Eds.; ASTM International: West Conshohocken, PA, USA, 2010; pp. 64–93, ISBN 978-0-8031-8417-6.

27. Lee, G.-G.; Kim, M.-C.; Lee, B.-S. Machine learning modeling of irradiation embrittlement in low alloy steel of nuclear power plants. *Nucl. Eng. Technol.* **2021**, *53*, 4022–4032. [CrossRef]
28. Kirk, M. A wide-range embrittlement trend curve for western reactor pressure vessel steels. *ASTM Spec. Tech. Publ.* **2013**, *1547*, 20–51. [CrossRef]
29. U.S. NRC. *Adjunct for ASTM E900-15: Technical Basis for the Equation used to Predict Radiation-Induced Transition Temperature Shift in Reactor Vessel Materials*; U.S. NRC: Washington, DC, USA, 2015.
30. Ferreño, D.; Sainz-Aja, J.A.; Carrascal, I.A.; Cuartas, M.; Pombo, J.; Casado, J.A.; Diego, S. Prediction of mechanical properties of rail pads under in-service conditions through machine learning algorithms. *Adv. Eng. Softw.* **2021**, *151*, 102927. [CrossRef]
31. Todeschini, P.; Kirk, M. Further assessment of the ASTM E900-15 transition temperature shift relationship. In Proceedings of the IGRDM-19: 19th Meeting of the International Group on Radiation Damage Mechanisms in Pressure Vessel Steels, Asheville, NC, USA, 10–15 April 2016.
32. Kirk, M.; Hashimoto, Y.; Nomoto, A.; Yamamoto, M.; Soneda, N. Application of a Machine Learning Approach Based on Nearest Neighbors to Extract Embrittlement Trends from RPV Surveillance Data. In Proceedings of the 2021 Meeting of the International Group on Radiation Damage, Mol, Belgium, 8–10 September 2021.