Conformal prediction for hypersonic flight vehicle classification

Zepu Xi XIZP3@MAIL2.SYSU.EDU.CN
Xuebin Zhuang ZHUANGXB@MAIL.SYSU.EDU.CN
Hongbo Chen CHENHONGBO@MAIL.SYSU.EDU.CN
School of Systems Science and Engineering, Sun Yat-sen University, Guangzhou, China

Editor: Henrik Boström, Lars Carlsson, Ulf Johansson, Zhiyuan Luo and Khuong An Nguyen

Abstract

This paper introduces a probabilistic guaranteed prediction method for trajectory data of the hypersonic flight vehicle classification problem. This paper devoted two problems: (1) hypersonic flight vehicle trajectory classification algorithm using functional data analysis method, and (2) a distributions-free uncertainty quantity for the classification results applying conformal prediction methodology. Our approach provides explicit finite-sample guarantees for any data set by using functional data analysis methods, which map the original data into feature space. The distribution-free uncertainty quantity results for the label of new objects include two indications, such as confidence and credibility respectively. Lastly, the proposed method aims to communicate instance-wise uncertainty under the probabilistic guaranteed and generate a prediction set at a user-specified confidence level for the hypersonic flight vehicle classification problem.

Keywords: Hypersonic flight vehicle; Conformal prediction; Distribution-free uncertainty Quantity; Functional data analysis; Support Vector Machines

1. Introduction

Making an accurate classification, based on the observed trajectories facts, is of intense demand in hypersonic flight vehicle (HFV) trajectory discrimination. Intuitively, in many respects, the mainstream of the classical point view of the trajectory is the time series analysis, and there is much more research using deep learning-based approaches to the raw observed trajectory for prediction and classification (Gaiduchenko and Gritsyk, 2019; Shi and Wang, 2020; Bartusiak et al., 2021; Xie et al., 2021). All of these reveal that the progress of classification has made it possible to solve ever more difficult problems in a wide variety of HFV classification fields. However, a major drawback remains that there is no confidence measure for the prediction of output value for particular new inputs (Gammerman and Vovk, 2006). For example, traditional machine learning and deep learning algorithms for pattern recognition just output simple predictions, without any associated uncertainty quantity for the results of algorithms (Vovk et al., 2005; Gawlikowski et al., 2021). Confidence and credibility measures of uncertainty quantity is an attraction indication of how likely each prediction is of being correct and it makes sense of information to evaluate the performance.

The prediction that is associated with some kind of valid uncertainty quantity is highly desirable in many risk-sensitive applications, such as those used for medical diagnosis (Nouretdinov et al., 2013), security (Balasubramanian et al., 2014), healthcare (Vazquez and Facelli, 2022), and various empirical sciences. In fact, such valid uncertainty quantity

information can benefit the application of the design adaptive systems, as the results of confidence values can be used to determine the way in which prediction will be arranged. For instance, a filtering discriminate system can be employed so that only predictions which satisfy a specified level of confidence will be taken into account. There are two main areas in mainstream learning algorithms that can be used in order to obtain some confidence measures: the Bayesian framework and the theory of probably approximately correct (PAC theory) (Valiant, 1984). In order to apply the Bayesian framework one is required to have some prior distribution knowledge about the data generation. When some prior knowledge on the distribution of the data is known, Bayesian methods provide valid optimal decisions. However, for empirical science, the required knowledge is not available. Generally, one has to assume the existence of a general "rule".

There is a commendable place that conformal prediction is a suitable and informative method to overcome this gap. Conformal prediction obtains confidence measure was proposed by Vovk et al. (2005). It is developed on top of any learning algorithm, especially for the "Block Box" models, to build distribution-free uncertainty quantity values. Instead of the traditional statistical methods, which provide the average treatment effect measurements, conformal prediction provides a measure of confidence and credibility measurements for individual treatment effects and does not require a specific distribution assumption. Confidence measures how likely our prediction is under the adoption of learning algorithms and the credibility serves as an indication of how suitable the training data are for classifying the test examples. In contrast with the Bayesian method and PAC theory, conformal prediction gives a probabilistic guarantee valid uncertainty measure, as they are only based on the general independently and identically distributed (I.I.D.) assumption.

Meanwhile, when we operate the hypersonic flight vehicle classification problem, data should be considered as discretized functions rather than as standard scale vectors. So from the point of the functional perspective, it provides a powerful tool for many empirical science processes. Typical examples arise out of temporal-spatial structure data, longitudinal data, and trajectory data (Rossi and Villa, 2006; Hyndman and Ullah, 2007; Martin-Barragan et al., 2014). The raw observation facts are the dependent data, as mentioned in Dashevskiy and Luo (2011); Chernozhukov et al. (2018); Xu and Xie (2021), to apply conformal prediction methods to dependent data, a major challenge is incompatible with the I.I.D. assumption. However, when we map the dependent data onto the feature space, represented by orthogonal basis function, and then we can apply the machine learning methods to the obtained finite-dimensional data, a conformal prediction approach to explore functional data with cluster can be found in Lei et al. (2015).

In this paper, we proposed a conformal prediction method of the underlying algorithm under functional data analysis to provide a probabilistic guarantee valid confidence and reliability set prediction for HFV classification. Using proposed conformal prediction methods, we can obtain the p-values, confidence, and reliability measure, and produce a set predictions for instance-wise examples to any specified significance level. Additionally, the validity and efficiency of learning algorithms under different functional data analysis methods, such as B-spline and the Fourier basis function methods, are introduced and the different significance levels results are demonstrated and evaluated on the HFV data sets by our methods. To our best knowledge, this is the first paper to do a set of prediction for HFV classification with finite sample probabilistic guarantees.

The paper is organized as follows: In Section 2 we introduce the dynamic models of HFV for simulation trajectory data. In Section 3.1 we introduce the functional data analysis tools for trajectory data representative. In Section 3.2 we introduce the Support Vector Machines as the underlying algorithms. In Section 3.3 we present the main theoretical properties of the proposed extended conformal prediction. In Section 4 we apply the proposed algorithm to efficiently construct a probabilistic guarantee set prediction for HFV trajectory functional data, the results are based on good finite sample property under any specified significance level and in Section 5 we give some conclusions.

2. Dynamic models of HFVs

Following Li et al. (2015), equations describing the motion of HFVs will be very complex when the rotation and non-spherical perturbations of the Earth are considered. But the maneuver characteristics of HFVs are little affected by the presence of mentioned effects, motion equations of HFVs can be deduced based on the assumption that the Earth is a homogeneous sphere without rotation and the sideslip angle is zero, and are of the form

$$\begin{aligned}
\dot{V} &= -D - g\sin\theta,\\
\dot{\theta} &= [L\cos\nu + (V^2/r - g)\cos\theta]/V,\\
\dot{\sigma} &= L\sin\nu/(V\cos\theta) + V\tan\phi\cos\theta\sin\sigma/r,\\
\dot{r} &= V\sin\theta,\\
\dot{\lambda} &= -V\cos\theta\sin\sigma/(r\cos\phi),\\
\dot{\phi} &= V\cos\theta\cos\sigma/r,
\end{aligned}$$

where $V, \theta, \sigma, r, \lambda, \phi$ are speed, flight path angle, velocity heading angle measured from North in a clockwise direction, the radial distance from the center of the Earth to the vehicle, longitude, and latitude, respectively. ν is bank angle, one of the two control variables. g is gravity acceleration. The lift and drag accelerations are defined as

$$\begin{aligned}
L &= \frac{1}{2}C_L\rho V^2 S_M/m,\\
D &= \frac{1}{2}C_D\rho V^2 S_M/m,
\end{aligned}$$

where $C_L = C_L(\alpha, M_\alpha)$ and $C_D = C_D(\alpha, M_\alpha)$ are lift and drag coefficients that are functions of angle of attack, α, and Mach number, Ma. S_M is the reference area and m is the vehicle mass. The atmospheric density is modeled by the following exponential function of altitude

$$\rho = \rho_0 \exp(-h/h_s),$$

where ρ_0 is the atmospheric density at sea level, and h_s is the scale-altitude. $h = r - R_e$ is the altitude and R_e is the radius of the Earth.

Generally, the pairs (r, λ, ϕ) denote the position of the vehicle in the space. In this paper we devoted the original observation $(r_n, \lambda_n, \phi_n) \in \mathbb{R}^3$ into position information $P_n(t)$ and then summary the position information by local-scale transform into functional data $X_n(t)$,

$$\begin{aligned}
P_n(t) &= \sqrt{r_n^2(t) + \lambda_n^2(t) + \phi_n^2(t)},\\
X_n(t) &= \log(P_n(t)) - \log(P_n(0)).
\end{aligned}$$

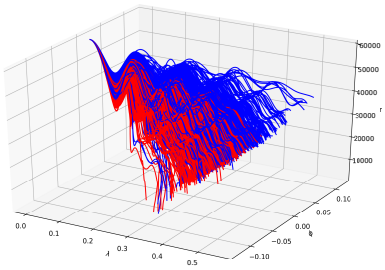

Figure 1: The three-dimensional trajectory plots of different maneuver models of HFV. Red lines: CAV-H, blue lines: CAV-L.

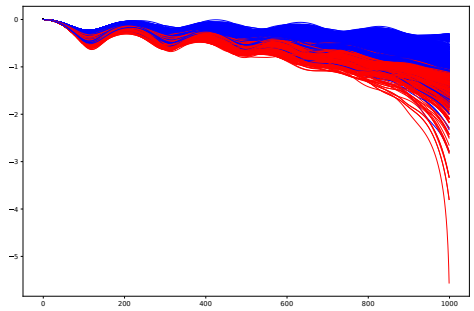

Figure 2: The local-scale transform data for three-dimensional trajectory plots of different maneuver models of HFV. The horizontal coordinate axis represents the observation time and the vertical coordinate axis represents the values of local-scale transform.

where n is the index of the HFV trajectory curve, and t is the observation time. We simulate two types of HFV, the CAV-H (indicated by -1) and CAV-L (indicated by 1), under different complex conditions. Figure 1 shows the three-dimensional trajectory plots of different maneuver models of HFV, the red lines are CAV-H, and the blue lines are CAV-L. Figure 2 show the local-scale transform data for three-dimensional trajectory. In addition, we apply functional data analysis tools to operate the local-scale transformed data to project the curves onto a finite-dimensional space feature space that captures the main features of the curves. Finally, the problem of trajectory data classification reduces to the problem of curve classification. We will demonstrate the details in Section 3.1.

3. Theory Analysis

3.1. Functional data analysis

In this section, we will give a sketches description of functional data analysis. The use of classical models for the raw observation facts face several difficulties: as the inputs are discretized functions, they are generally represented by high-dimensional vectors whose coordinate are highly dependent. As a consequence, many traditional statistical methods will lead to ill-posed problems, both from a theoretical point of view and on a practical one. More details of the anatomy of functional data analysis can be found (Ramsay et al., 2009) and (Kokoszka and Reimherr, 2017), which have established the state-of-the-art instruction.

A common approach to functional data analysis is mapping the raw data into feature space that compresses the main features of the curves. The simplest data set encountered with functional data analysis is a sample of the form

$$X_n(t) \in \mathbb{R}, n = 1, 2, \ldots, N, t = 1, \ldots, T, \tag{1}$$

where N is the number of curves and T is the observation duration.

For the simplified, the developments of this paper in working with functional data onto the form (1) are to express them by means of a basis expansion

$$X_n(t) = \sum_{m=1}^{M} c_{nm} B_m(t), 1 \leq n \leq N, \tag{2}$$

where B_m is some standard collection of basis functions, like the B-splines, the wavelets, or the Fourier basis functions and c_{nm} are the estimated coefficients. Expansion (2) reflects the intuition that the observation facts are smooth functions that share the same shape properties, and so can be approximated as linear combinations of some M basis shapes B_m, with M being typically small than the number of observed time duration, T. If the observation duration T of curves is very large, as far $T = 1000s$ in this paper, expansion (2) serves the practical purpose of replacing the original scalar data $X_n(t)$ by a smaller collection of the coefficients c_{nm}. For each curve $X_n(t)$ is represented by the column vector $c_n = [c_{n1}, c_{n2}, \ldots, c_{nM}]^T$ of dimension M. Additional one of the most useful tools of functional data analysis is the estimation of principal component analysis. In expansion (2), the basis function B_m is fixed. The idea of the functional principal component expansion is to find function $\hat{v}_j$ such that the smoothing function $X_n(t) = \sum_{m=1}^{M} c_{nm} B_m(t)$ is represented as

$$X_n(t) = \sum_{m=1}^{M} c_{nm} B_m(t) = \sum_{j=1}^{p} \hat{\epsilon}_{nj} \hat{v}_j(t), \tag{3}$$

with the number of component p much smaller than M in (3). The estimated functional principal components, EFPC's $\hat{v}_j$, are computed from the observed function $X_1, X_2, \ldots, X_N$ after converted them into functional objects.

In classification setting, the each curve $X_n(t)$ is represented by EFPC's vector $\hat{\epsilon}_n = [\hat{\epsilon}_{n1}, \hat{\epsilon}_{n2}, ..., \hat{\epsilon}_{np}]^T$, while p being typically small than the number of basis M. We will build pattern recognition algorithms based functional data and develop the uncertainty quantity of the prediction.

3.2. Support Vector Machines

We give, in this section, a very brief presentation of SVMs that is needed for the classification task. We refer the reader to Vapnik (1995, 1998, 2006) for a more comprehensive presentation. In this section, we will assume, for simplicity, that our classification problem is binary, and the possible labels are -1 and 1. Our goal is to classify data into two predefined classes when the training data are nonseparable (Cortes and Vapnik, 1995).

Let the set of training set

$$(x_1, y_1), \ldots, (x_l, y_l), x \in \mathbb{R}^n, y \in \{-1, 1\},$$

be such that it cannot be separated without error by a hyperplane. Our goal is to construct the hyperplane that makes the smallest number of error. To get a formal setting of this problem we introduce the nonnegative variables

$$\xi_1, \ldots, \xi_l.$$

Thus, we will minimize the functional

$$R(w, b) = \sum_{i=1}^{l} \xi_i, \tag{4}$$

subject to the constraints

$$y_i((w, z_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \ldots, l, \tag{5}$$

and the constraint

$$(w, w) \leq h. \tag{6}$$

where the VC dimension h of the set of hyperplanes. z is the image apace of input x with the operator $z = \mathcal{F}x$.

To solve this optimization problem we find the saddle point of the Lagrangian

$$L(w, b, \alpha, \beta, \gamma) = \sum_{i=1}^{l} \xi_i - \frac{1}{2}\gamma((w, w) - h) - \sum_{i=1}^{l} \alpha_i(y_i(w, w) + b) - 1 + \xi_i) - \sum_{i=1}^{l} \beta_i \xi_i, \tag{7}$$

(the minimum with respect to w, b, ξ_i and the maximum with respect to nonnegative multipliers $\alpha_i, \beta_i, \gamma$). The parameters that minimize the Lagrangian must satisfy the conditions

$$\begin{aligned}
\frac{\partial L(w, b, \alpha, \beta, \gamma)}{\partial w} &= \gamma w - \sum_{i=1}^{l} \alpha_i y_i z_i = 0, \\
\frac{\partial L(w, b, \alpha, \beta, \gamma)}{\partial b} &= -\sum_{i=1}^{l} y_i \alpha_i = 0, \\
\frac{\partial L(w, b, \alpha, \beta, \gamma)}{\partial \xi_i} &= 1 - \alpha_i - \beta_i = 0.
\end{aligned}$$

From these conditions one derives

$$w = \frac{1}{\gamma}\sum_{i=1}^{l}\alpha_i y_i z_i, \tag{8}$$

$$\sum_{i=1}^{l}\alpha_i y_i = 0, \tag{9}$$

$$\alpha_i + \beta_i = 1. \tag{10}$$

Substituting (8) into the Lagrangian and taking into account (9,10), we obtain the functional

$$W(\alpha) = \sum_{i=1}^{l}\alpha_i - h\sqrt{\sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j (z_i, z_j)}, \tag{11}$$

subject to the constraint

$$\sum_{i=1}^{l} y_i \alpha_i = 0, \tag{12}$$

and the constraints

$$0 \leq \alpha_i \leq 1, i = 1, \ldots, l.$$

3.3. Conformal prediction

We give, in this section, a brief introduction to the main idea about conformal prediction. In clarifying the origin of the conformal prediction, the more details see Vovk et al. (2005); Gammerman and Vovk (2006); Shafer and Vovk (2008). The successive sequence pairs

$$z_1, \ldots, z_l,$$

are assumed to be generated independently from the same probability distribution, and $z_i = (x_i, y_i)$ are called examples. Each example z_i consists of an object x_i and its label y_i. We are also given a new candidate example x_{l+1} and our task is to predict the classification y_{l+1} of this candidate example. Our standard assumption is that all the problems we are interested in are the general I.I.D. and we know a prior of the set of all possible labels $Y_1, \ldots, Y_c$. Now suppose we can test how likely it is that a successive sequence of classified examples was randomness; in other words how typical the sequence is respected to the I.I.D. model. From the point of view of the algorithmic theory of randomness, we can make a confidence set prediction for the label y_{l+1} of the new object x_{l+1} given training set $z_1, \ldots, z_l$ if the algorithmic randomness deficiency is small for only one possible extension

$$(z_1, \ldots, z_l, (x_{l+1}, y)), y \in Y_1, \ldots, Y_c,$$

we can then output the corresponding y as a confident set prediction. The reference to the algorithmic theory of randomness can consult Kolmogorov (1963); Martin-Löf (1966); Kolmogorov (1968, 1983); Li and Vitányi (2008). In the spirit of Martin-Löf (1966), a function $p : Z^* \to [0, 1]$ is a test for randomness with respect to the I.I.D. model if

1. $\forall n \in \mathbb{N}, \forall \delta \in [0,1]$ and for all probability distribution P on Z,

$$P^n\{z \in Z^n : p(z) \leq \delta\} \leq \delta, \tag{13}$$

2. p is semi-computable above.

In the statistics community, this randomness test definition is generally equivalent to the notion of p-values. Therefore, we can obtain a randomness test measure of a successive sequence by using the computed p-value of label $Y_j, j = 1, \ldots, c$ and denote it by $p(Y_j)$.

As shown in Section 3.2 and Vovk et al. (2005), any conformal prediction is exactly valid, that means will produce independent p-values $p(Y_j)$ distributed uniformly in $[0,1]$. The proof of the theorem can be found in Vovk et al. (2005).

In practice, if the p-value of a given label is under the significance level, say 0.05, this means that the sequence

$$((x_1, y_1), \ldots, (x_l, y_l), (x_{l+1}, Y_j)),$$

would reject the null hypothesis, say the sequence was drawn I.I.D. model. To compute how different each example to the successive sequence is from with all other examples, we use a family of functions $A_n : Z^{(n-1)} \times Z \to \mathbb{R}, n = 1, 2, \ldots$, which assign a numerical score of each example z_i,

$$\alpha_i = A_n((z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_l), z_i), \tag{14}$$

which indicate how different it is from the examples of the classified $(z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_l)$; such families of measure functions are called nonconformity measures function. For a new test object x_{l+1}, as we know the prior of label $Y_j, j = 1, \ldots, c$, we can compute the prediction values under each labels with the aid of some learning algorithm, which we call the underlying algorithm D,

$$y_{l+1}^{Y_j} = D(x_{l+1}), j = 1, \ldots, c. \tag{15}$$

After that, we can measure how different the new prediction $y_{l+1}^{Y_j}$ is from the examples in the training sequences by ranking their dissimilarity, such as

$$p(Y_j) = \frac{\#\{i = 1, \ldots, l, l+1 : \alpha_i^{(Y_j)} \geq \alpha_{l+1}^{(Y_j)}\}}{l+1}, j = 1, \ldots, c. \tag{16}$$

When we obtain the p-value, $p(Y_j), j = 1, \ldots, c$, of prediction of the new object x_{l+1}, embedding the prediction label with the largest p-value with credibility, and output confidence to this prediction one minus the second-largest p-value.

For Inductive Conformal Prediction, in this paper, we consider SVMs and other learner as the underlying algorithm of ICP. The empirical data $z_i = (x_i, y_i), i = 1, 2, \ldots, g$, it is first randomly split into two sets, training set $Z_{\textbf{training}} = \{z_1, \ldots, z_l\}$ and test set $Z_{\textbf{test}} = \{z_{1+1}, \ldots, z_g\}$. The training set is further divided into the proper training set $Z_{\textbf{proper}} = \{z_1, \ldots, z_m\}$ and calibration set $Z_{\textbf{calibration}} = \{z_{m+1}, \ldots, z_l\}$. The learning model is trained on the proper training set $Z_{\textbf{proper}}$ with default parameters and then nonconformity

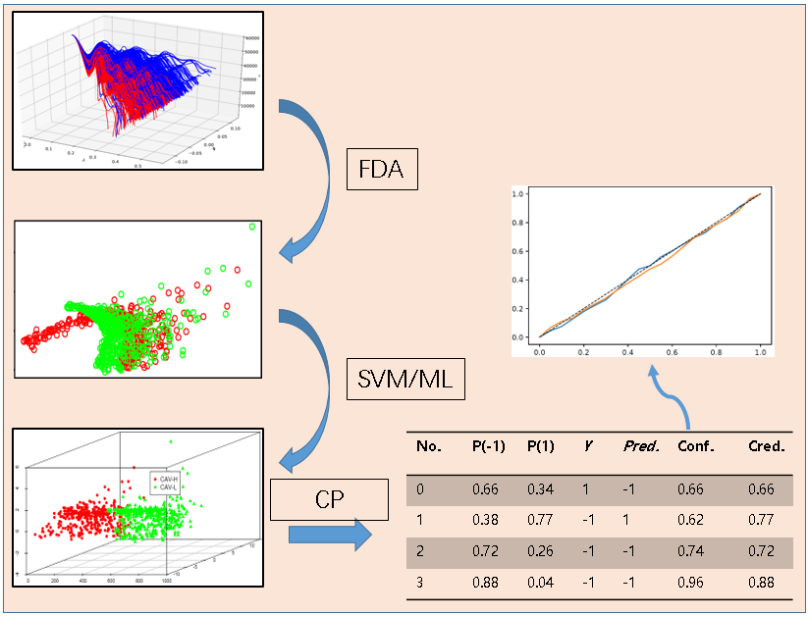

Figure 3: The general flow of the proposed algorithm.

scores, p-values, and conformal predictors will be computed upon calibration and test tests, respectively.

For an example of calibration sets $z_{m+i} \in Z_{\text{calibration}}$, algorithms outputs probability values $y_{l+1}^{Y_j}$ and its nonconformity score is computed by Eq. (14). For the test examples, the predicted probabilities are transducted by underlying algorithm and the nonconformity scores and p-values are calculated by Eq. (14) and Eq. (16) respectively.

As we are mainly interested in the confidence levels of 99%, 95%, 90%, and 80%, we can compare the value of p-value under the different confidence levels, and obtain the probabilistic guarantee classification. From all of the descriptions above, we summarize the general steps of the proposed algorithm a flow graphic as the Figure 3.

4. Results

Here we detail the experimental results of conformal prediction on the hypersonic flight vehicle trajectories data sets, in which the dynamic model was introduced in Section 2. We simulate 2400 trajectories for each maneuver model, constructing functional data analysis object of B-spline basis function, the number of estimated principle components p are computed from the observed functions $X_1, X_2, \ldots, X_N$ after converting them into functional objects. Finally, we split the feature data into the proper training data, calibration data, and test data according to 2:1:1, and build classification algorithms based on the proper training data in feature space.

We compute the nonconformity α_i of each example $z_i = (x_i, y_i)$ in successive sequence $(z_1, \ldots, z_l)$ with the underlying algorithms, such as SVMs, Boosting and Neural Networks. The underlying algorithm creates a prediction decision rule $D_{(z_1,\ldots,z_l)}$, which maps any unlabeled candidate example x_{l+1} to a prediction label $\hat{y}_{l+1}$. As during the training processes,

the decision rule is built to the examples in the given sequences, the prediction labeled $\hat{y}_i = D(x_i)$, which tell us how analogous z_i is from the rest of the examples in the sequence, the higher the output $\hat{y}_i$ corresponding to the label, the more conforming the example. Therefore, this dissimilarity α_i can be measured as *inverse probability* (Aleksandrova and Chertov, 2021),

$$\alpha_i = -\hat{y}_i, \tag{17}$$

which gives us a measure of the nonconformity of example z_i. Our experiments on these data sets were performed using the SVMs model, with the default parameter, such as regularization parameter C=1.0, radial basis function kernel, and RBF kernel coefficient γ=0.1. Here we report the accuracy percentage of SVMs with the conformal prediction (SVMs-CP) and compare them to the ones of its underlying algorithm as well as those other learning methods, such as decision tree (DT), boosting, neural network (NN), naïve Bayes (NB), and other classical statistics models, such as logistic regression (LR) and linear discriminant analysis (LDA). All of these experiments are implements with an open-source `Python` package $\texttt{scikit} - \texttt{learn}$ based on the default parameters settings. The addition of the results that we compare the prediction performance under the different basis functions at functional data analysis, such as adding the Fourier basis function. Table 1 lists the results, which show that the prediction methods have the improvement to the classical statistics model and there is a less significant performance for the B-spline and the Fourier basis function.

For practice, along with the accuracy, we also check the quality of its p-values by analyzing the results for the 99%, 95%, 90%, and 80% confidence levels. For the purpose of reporting the results, we separate its output into four categories:

1. $> 1(\%)$: The percentage that the prediction set with more than one labels.
2. $= 1(\%)$: The percentage that the prediction set only one label.
3. $\emptyset(\%)$: The percentage that the prediction set is empty set.
4. Accuracy(%): The percentage of correct classification.

Our main concern here will be the number of outputs that belong to the first category, that the prediction set with more than one label, we want this number to be relatively small since there are the indicates for which the conformal prediction is not certain in only one label at the specified confidence level $1 - \epsilon$. In addition to the percentage of examples of each category, we also report the prediction set only one label and the number of accuracy made by conformal prediction in underlying algorithms. Finally, we report all the cases where the prediction set was empty, we want this number to be converged on zero since there is supported evidence that the underlying algorithm makes nonsense for classification tasks. All the information is listed in Table 2. If the empirical accuracy is approximate to the theory suggestion, which is defined by confidence level $1 - \epsilon$, we conclude that the prediction is validity.

To confirm the validity of confidence measures, we check the empirical distribution of each category p-value for different underlying algorithms, as shown in Figure 4. The

Table 1: Accuracy rate comparison of the SVMs-CP with traditional algorithms.

Underlying algorithm	Accuracy(%)	
	B-spline	Fourier
SVMs	74.71	74.71
DT	75.37	72.05
Boosting	73.38	74.38
NN	72.55	74.04
NB	58.74	57.07
LR	46.59	46.92
LDA	46.59	46.92

results support the evidence that the proposed probabilistic guarantee method is valid. The demonstration of uncertainty measures are the list in Table 3, the first column is sample index, the rest are each category p-value, true label, conformal prediction label (CP-Label), confidence, and credibility uncertainty measure respectively. Surely that we can produce the instance-wise uncertainty measures for the SVMs algorithm and exact more information about each example than a simple prediction, which usually just takes the mean square measure for test examples. Table 4 illustration selected test examples from the HFV data set, and show the prediction set at a user-specified significance level. It contains information on the p-values of labels (-1 and 1), significance levels, and true labels respectively.

Table 2: Results of the B-spline basis function of conformal prediction for HFV data sets.

Algorithm	1-ϵ	Accuracy(%)	=1(%)	>1(%)	$\emptyset$(%)
SVMs	99%	99.66	18.14	81.86	0.00
	95%	94.16	40.10	59.90	0.00
	90%	90.65	52.25	47.75	0.00
	80%	81.70	79.03	20.97	0.00
DT	99%	99.67	27.45	72.54	0.00
	95%	96.17	44.93	55.07	0.00
	90%	92.20	57.90	42.10	0.00
	80%	82.68	82.03	18.00	0.00
Boosting	99%	99.83	8.49	91.51	0.00
	95%	96.88	25.96	70.04	0.00
	90%	91.50	49.25	50.75	0.00
	80%	82.89	76.71	23.29	0.00
NN	99%	100	7.82	92.17	0.00
	95%	95.32	34.44	65.56	0.00
	90%	92.12	46.42	53.58	0.00
	80%	80.50	81.70	18.30	0.00
NB	99%	99.49	4.50	95.50	0.00
	95%	96.00	15.14	84.85	0.00
	90%	92.66	24.63	75.37	0.00
	80%	80.56	47.59	52.41	0.00
LR	99%	99.67	4.66	95.34	0.00
	95%	94.31	15.31	84.69	0.00
	90%	89.17	22.96	77.04	0.00
	80%	78.80	39.10	60.89	0.00
LDA	99%	99.67	4.66	95.34	0.00
	95%	94.31	15.31	84.69	0.00
	90%	89.17	22.96	77.04	0.00
	80%	78.80	39.10	60.89	0.00

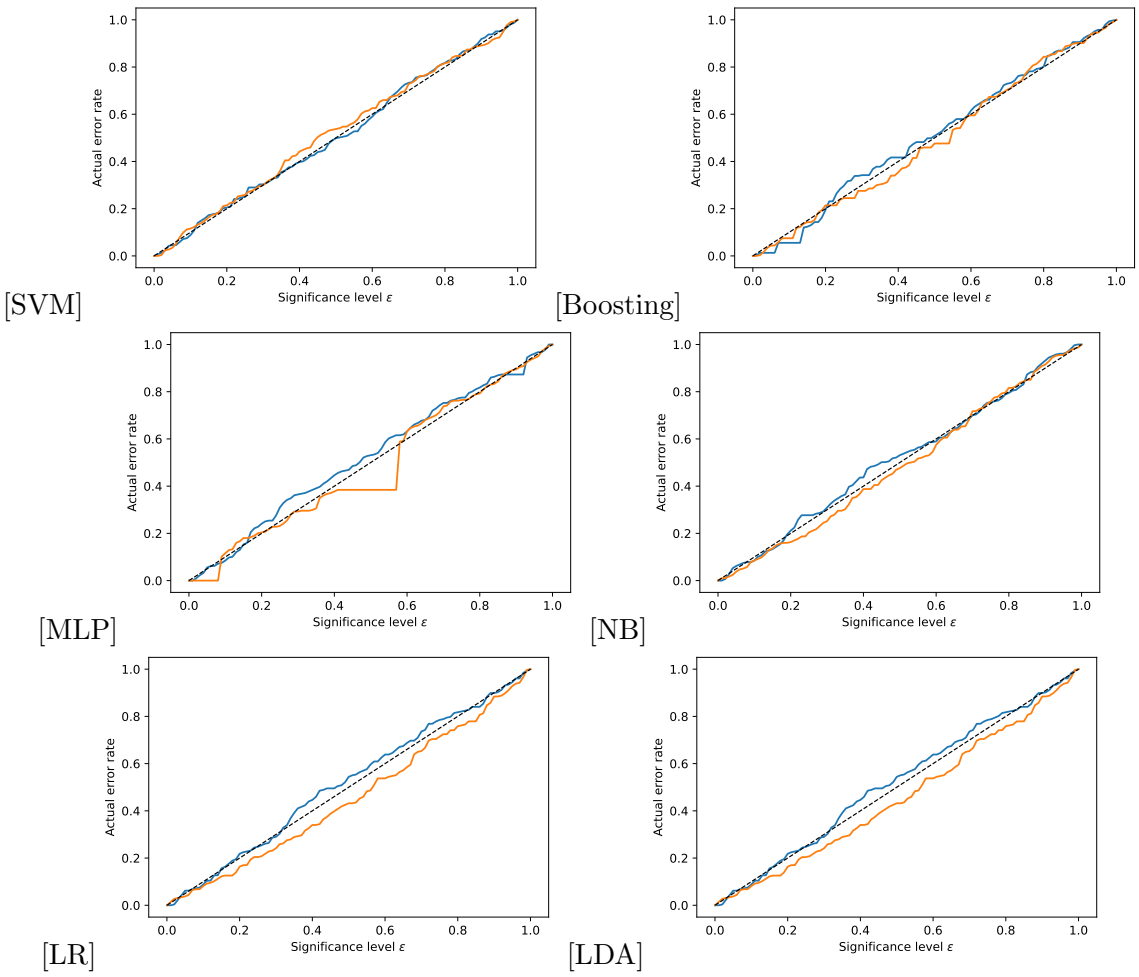

Figure 4: The empirical distribution of each category p-values for different underlying algorithms.

Table 3: The demonstration results of conformal prediction method underlying SVMs algorithm.

#	-1	1	True Lable	CP-Label	Confidence	Credibility
0	0.872483	0.613115	-1	-1	0.386885	0.872483
1	0.845638	0.619672	-1	-1	0.380328	0.845638
2	0.184564	0.960656	1	1	0.815436	0.960656
3	0.825503	0.655738	-1	-1	0.344262	0.825503
4	0.755034	0.718033	-1	-1	0.281967	0.755034
5	0.446309	0.849180	1	1	0.553691	0.849180
6	0.580537	0.777049	1	1	0.419463	0.777049
7	0.711409	0.724590	1	1	0.288591	0.724590
8	0.963087	0.419672	-1	-1	0.580328	0.963087
9	0.436242	0.852459	1	1	0.563758	0.852459
10	0.718121	0.721311	-1	1	0.281879	0.721311
11	0.718121	0.721311	-1	1	0.281879	0.721311
12	0.718121	0.721311	-1	1	0.281879	0.721311

Table 4: The demonstration prediction set of conformal prediction method underlying SVMs algorithm.

#	-1	1	0.01	0.05	0.1	0.15	0.2	0.25	0.5	0.75	0.8	0.85	0.9	0.95	0.99	1	True Label
1	0.382550	0.773770	[-1, 1]	[-1, 1]	[-1, 1]	[-1, 1]	[-1, 1]	[-1, 1]	[1]	[1]	[]	[]	[]	[]	[]	[]	-1
2	0.721477	0.245902	[-1, 1]	[-1, 1]	[-1, 1]	[-1, 1]	[-1, 1]	[-1]	[-1]	[]	[]	[]	[]	[]	[]	[]	-1
3	0.882550	0.036066	[-1, 1]	[-1]	[-1]	[-1]	[-1]	[-1]	[-1]	[-1]	[-1]	[-1]	[]	[]	[]	[]	-1
4	0.053691	0.940984	[-1, 1]	[-1, 1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[]	[]	[]	-1
5	0.120805	0.901639	[-1, 1]	[-1, 1]	[-1, 1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[]	[]	[]	-1
6	0.087248	0.934426	[-1, 1]	[-1, 1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[]	[]	[]	-1
7	0.174497	0.875410	[-1, 1]	[-1, 1]	[-1, 1]	[-1, 1]	[1]	[1]	[1]	[1]	[1]	[1]	[]	[]	[]	[]	-1
8	0.013423	0.993443	[-1, 1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[]	1
9	0.194631	0.865574	[-1, 1]	[-1, 1]	[-1, 1]	[-1, 1]	[1]	[1]	[1]	[1]	[1]	[1]	[]	[]	[]	[]	-1

5. Conclusions

This paper presents a probabilistic guarantee method of hypersonic flight vehicle classification problem and produces instance-wise uncertainty measures with confidence and credibility for each test example under the distribution-free framework. First, we introduce the functional data analysis methods and mapping the raw observation facts into feature space, and then build the SVMs method to achieve the classification results. Moreover, we introduce conformal prediction methods to produce the uncertainty quantity of the algorithm outputs. All the measurement results are based on the underlying algorithm and data distribution assumptions are not required. The proposed method accompanies each of their predictions with probabilistic guarantee valid measures of confidence.

For the functional data analysis, we also produce two basis functions, B-spline and the Fourier basis function, and compare the classification performance, that there is less significant prediction accuracy improvement. Different from the traditional simple prediction method, our new method provides the instance-wise uncertainty measure for any underlying algorithms in any specified significance levels. In addition, that the aim of the proposed methods is not just to compare the accuracy with simple prediction but more information with each example.

References

Zainab Abohashima, Mohamed Elhosen, Essam H. Houssein, and Waleed M. Mohamed. Classification with quantum machine learning: A survey. *arXiv:2006.12270*, 2020.

Marharyta Aleksandrova and Oleg Chertov. Impact of model-agnostic nonconformity functions on efficiency of conformal classifiers: an extensive study. *Proceedings of Machine Learning Research*, 152, 2021.

Vineeth Balasubramanian, Shen-Shyang Ho, and Vladimir Vovk. Conformal prediction for reliable machine learning. *Morgan Kaufmann*, 2014.

Emily R. Bartusiak, Nhat X. Nguyen, Moses W. Chan, Mary L. Comer, and Edward J. Delp. A machine learning approach to classify hypersonic vehicle trajectories. *IEEE*, March 2021. doi: 10.1109/aero50100.2021.9438274.

Luc Bégin, Pascal Germain, François Laviolette, and Jean-Francis Roy. PAC-Bayesian Theory for Transductive Learning. *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, pages 105–113, 2014.

José Cabrero-Holgueras and Sergio Pastrana. SoK: Privacy-preserving computation techniques for deep learning. *Proceedings on Privacy Enhancing Technologies*, 2021(4):139–162, 2021.

Chuan Chen, Rui Li, Lin Shu, Zhiyu He, Jining Wang, Chengming Zhang, Huanfei Ma, Kazuyuki Aihara, and Luonan Chen. Predicting future dynamics from short-term time series using an anticipated learning machine. *National Science Review*, 7(6):1079–1091, February 2020. doi: 10.1093/nsr/nwaa025.

Victor Chernozhukov, Kaspar Wuthrich, and Yinchu Zhu. Exact and robust conformal inference methods for predictive machine learning with dependent data. *arXiv:1802.06300*, 2018.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3): 273–297, September 1995. doi: 10.1007/bf00994018.

M. Dashevskiy and Z. Luo. Time series prediction with performance guarantee. *IET Communications*, 5(8):1044–1051, May 2011. doi: 10.1049/iet-com.2010.0121.

Nikolai E. Gaiduchenko and Pavel A. Gritsyk. Hypersonic vehicle trajectory classification using convolutional neural network. *2019 International Conference on Engineering and Telecommunication (EnT)*, November 2019. doi: 10.1109/ent47717.2019.9030537.

Alexander Gammerman and Vladimir Vovk. Hedging predictions in machine learning. *The Computer journal*, 50(2), 2006.

Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, Muhammad Shahzad, Wen Yang, Richard Bamler, and Xiao Xiang Zhu. A survey of uncertainty in deep neural networks. *arXiv:2107.03342*, 2021.

Andrew Gelman and Christian Hennig. Beyond subjective and objective in statistics. *journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(4):967–1033, August 2017. doi: 10.1111/rssa.12276.

Weiyuan Gong and Dong-Ling Deng. Universal adversarial examples and perturbations for quantum classifiers. *National Science Review*, July 2021. doi: 10.1093/nsr/nwab130.

Sander Greenland, Stephen J. Senn, Kenneth J. Rothman, John B. Carlin, Charles Poole, Steven N. Goodman, and Douglas G. Altman. Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. *European journal of Epidemiology*, 31(4):337–350, April 2016. doi: 10.1007/s10654-016-0149-3.

Rob J. Hyndman and Md. Shahid Ullah. Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, 51(10): 4942–4956, June 2007. doi: 10.1016/j.csda.2006.07.028.

Piotr Kokoszka and Matthew Reimherr. Introduction to functional data analysis. *Chapman and Hall/CRC*, September 2017. doi: 10.1201/9781315117416.

A.N. Kolmogorov. On the tables of random numbers. *The Indian journal of Statistics, Series A*, 25(4):369–376, 1963.

A.N. Kolmogorov. Three approaches to the quantitative definition of information. *International journal of Computer Mathematics*, 1-4(2):157–168, 1968.

A.N. Kolmogorov. Combinatorial foundations of information theory and the calculus of probabilities. *Russian Math. Surveys*, 38(4):27–36, 1983.

Jing Lei, Alessandro Rinaldo, and Larry Wasserman. A conformal prediction approach to explore functional data. *Annals of Mathematics and Artificial Intelligence*, 74(1-2):29–43, September 2015. doi: 10.1007/s10472-013-9366-6.

Guanghua Li, Hongbo Zhang, and Guojian Tang. Maneuver characteristics analysis for hypersonic glide vehicles. *Aerospace Science and Technology*, 43:321–328, June 2015. doi: 10.1016/j.ast.2015.03.016.

Ming Li and Paul Vitányi. An introduction to kolmogorov complexity and its applications. *Springer New York*, 2008.

Belen Martin-Barragan, Rosa Lillo, and Juan Romo. Interpretable support vector machines for functional data. *European journal of Operational Research*, 232(1):146–155, January 2014. doi: 10.1016/j.ejor.2012.08.017.

Per Martin-Löf. The definition of random sequences. *Information and Control*, 9(6):602–619, December 1966. doi: 10.1016/s0019-9958(66)80018-9.

Thomas Melluish, Craig Saunders, Ilia Nouretdinov, and Volodya Vovk. Comparing the bayes and typicalness frameworks. *Machine Learning: ECML 2001*, 2001.

Ilia Nouretdinov, Dmitry Devetyarov, Volodya Vovk, Brian Burford, Stephane Camuzeaux, Aleksandra Gentry-Maharaj, Ali Tiss, Celia Smith, Zhiyuan Luo, Alexey Chervonenkis, Rachel Hallett, Mike Waterfield, Rainer Cramer, John F. Timms, Ian Jacobs, Usha Menon, and Alex Gammerman. Multiprobabilistic prediction in early medical diagnoses. *Annals of Mathematics and Artificial Intelligence*, 74(1-2):203–222, July 2013. doi: 10.1007/s10472-013-9367-5.

Harris Papadopoulos. Inductive conformal prediction: Theory and application to neural networks. *Tools in Artificial Intelligence*, August 2008. doi: 10.5772/6078.

Harris Papadopoulos, Volodya Vovk, and Alex Gammermam. Conformal prediction with neural networks. *19th IEEE International Conference on Tools with Artificial Intelligence(ICTAI 2007)*, October 2007. doi: 10.1109/ictai.2007.47.

Harris Papadopoulos, Vladimir Vovk, and Alex Gammerman. Regression conformal prediction with nearest neighbours. *journal Of Artificial Intelligence Research*, 40:815–840, 2014. doi: 10.1613/jair.3198.

James Ramsay, Giles Hooker, and Spencer Graves. Functional data analysis with r and MATLAB. *Springer New York*, 2009.

Blake A. Richards, Timothy P. Lillicrap, Philippe Beaudoin, Yoshua Bengio, Rafal Bogacz, Amelia Christensen, Claudia Clopath, Rui Ponte Costa, Archy de Berker, Surya Ganguli, Colleen J. Gillon, Danijar Hafner, Adam Kepecs, Nikolaus Kriegeskorte, Peter Latham, Grace W. Lindsay, Kenneth D. Miller, Richard Naud, Christopher C. Pack, Panayiota Poirazi, Pieter Roelfsema, João Sacramento, Andrew Saxe, Benjamin Scellier, Anna C. Schapiro, Walter Senn, Greg Wayne, Daniel Yamins, Friedemann Zenke, Joel Zylberberg, Denis Therien, and Konrad P. Kording. A deep learning framework for neuroscience. *Nature Neuroscience*, 22(11):1761–1770, October 2019.

Fabrice Rossi and Nathalie Villa. Support vector machine for functional data classification. *Neurocomputing*, 69(7-9):730–742, March 2006. doi: 10.1016/j.neucom.2005.12.010.

Craig Saunders, Alexander Gammerman, and Vladimir Vovk. Transduction with confidence and credibility. *journal of Machine Learning Research*, 2:722–726, 1999.

Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *journal of Machine Learning Research*, 9(12):371–421, 2008.

Yang Shi and Zhenbo Wang. A deep learning-based approach to real-time trajectory optimization for hypersonic vehicles. *American Institute of Aeronautics and Astronautics*, January 2020.

L. G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, November 1984. ISSN 0001-0782. doi: 10.1145/1968.1972.

Vladimir Vapnik. The nature of statistical learning theory. *Springer-Verlag New York, Inc*, 1995.

Vladimir Vapnik. Statistical learning theory. *Wiley-Interscience*, 1998.

Vladimir Vapnik. Estimation of dependences based on empirical data. *Springer New York*, 2006. doi: 10.1007/0-387-34239-7.

Janette Vazquez and Julio C. Facelli. Conformal prediction in clinical medical sciences. *Journal of Healthcare Informatics Research*, 2022.

V. Vovk, A. Gammerman, and C. Saunders. Machine-learning applications of algorithmic randomness. In *Sixteenth International Conference on Machine Learning (ICML-1999) (01/01/99)*, pages 444–453, 1999.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. Algorithmic learning in a random world. *Springer Science & Business Media*, 2005.

Ling Wang. Mu-ming poo: China brain project and the future of chinese neuroscience. *National Science Review*, 4(2):258–263, February 2017. doi: 10.1093/nsr/nwx014.

Yangfan Xie, Xuebin Zhuang, Zepu Xi, and Hongbo Chen. Dual-channel and bidirectional neural network for hypersonic glide vehicle trajectory prediction. *IEEE Access*, 2021.

Chen Xu and Yao Xie. Conformal prediction interval for dynamic time-series. *Proceedings of Machine Learning Research*, 139:11559–11569, 2021.

Yu Yu and Xiang Xie. Privacy-preserving computation in the post-quantum era. *National Science Review*, July 2021. doi: 10.1093/nsr/nwab115.