ECCOMAS
Proceedia

SENSITIVITY ANALYSIS OF IMAGE CLASSIFICATION MODELS USING GENERALIZED POLYNOMIAL CHAOS

Lukas Bahr[1,2], Lucas Poßner[2,5], Konstantin Weise[3,4], Stephan Matzka[5], Sophie Gröger[6] and Rüdiger Daub[7,8]

[1] TUM School of Engineering and Design, Technical University of Munich, Munich, Germany.
e-mail: lukas.bahr@tum.de

[2] Digitization Department for Production, BMW Group, Germany.

[3] Institute for Electrical Power Engineering, Leipzig University of Applied Sciences, Leipzig, Germany.

[4] Methods and Development Group "Brain Networks", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany.

[5] Professorship for Mechatronics and Mechatronic Systems, HTW Berlin, Berlin, Germany.

[6] Professorship of Production Measurement Technology, Technical University of Chemnitz, Chemnitz, Germany.

[7] Institute for Machine Tools and Industrial Management, Technical University of Munich, Munich, Germany.

[8] Institute for Casting, Composite and Processing Technology (IGCV), Fraunhofer, Augsburg, Germany.

Abstract. *Integrating advanced communication protocols in production has accelerated the adoption of data-driven predictive quality methods, notably machine learning (ML) models. However, ML models in image classification often face significant uncertainties arising from model, data, and domain shifts. These uncertainties lead to overconfidence in the classification model's output. To better understand these models, sensitivity analysis can help to analyze the relative influence of input parameters on the output. This work investigates the sensitivity of image classification models used for predictive quality. We propose modeling the distributional domain shifts of inputs with random variables and quantifying their impact on the model's outputs using Sobol indices computed via generalized polynomial chaos (GPC). This approach is validated through a case study involving a welding defect classification problem, utilizing a fine-tuned ResNet18 model and an emblem classification model used in BMW Group production facilities.*

Keywords: Sensitivity Analysis, Generalized Polynomial Chaos, Image Classification, Data-Driven Predictive Quality.

1 INTRODUCTION

Predictive quality has prevailed in the manufacturing industry thanks to the approaches defined in ISO / IEC 31010 [1] and include approaches such as failure mode and effects analysis (FMEA) [2], or fault tree analysis [3]. Recent advancements in large language models have enhanced their ability to reason with symbolic data, thereby maintaining the importance of these methods for risk assessment [4]. The rise of the industrial internet of things has enhanced operational efficiency by integrating advanced communication protocols. This development fosters a focus on data-driven predictive quality methods in production processes, such as the application of machine learning (ML) [5, 6]. Methods such as anomaly detection or classification algorithms can help improve product quality and reliability by finding patterns in data that are not immediately obvious to humans [7, 8]. In [9], the authors explore common approaches to ML that address predictive quality. Particular image classification models for predicting defects (nOK/OK), such as convolutional neural networks, are dominant in the manufacturing domain.

Understanding the sensitivity of image classification models is of great interest. Sensitivity analysis (SA) studies how the uncertainty of a model's output can be attributed to the uncertainty of its inputs [10]. This understanding is crucial because many models, especially those provided by external suppliers, function as black boxes, making it difficult to understand the inner workings. For instance, consider an image classification model that predicts the quality of a weld seam. Image classification models are often poorly calibrated and overconfident on the training data [11]. Thus, evaluating the sensitivity is essential to determine how changes in input, such as varying lighting conditions in the production environment, affect the model's reliability.

The challenges for analyzing the sensitivity can be summarized as follows [12, 13]:

1. Many ML models are black-box models, so we cannot explain the predictions given the inputs (model uncertainty).

2. Errors in the inputs that arise from, for example, inadequate measurements introduce uncertainty in the output (data uncertainty).

3. Distributional domain shifts in the inputs that arise from, e.g., changes in the production environment, introduce uncertainty in the output (domain uncertainty).

This paper investigates the sensitivity of image classification models to perturbations in the input data. We propose applying geometric transformations to the input images, such as changes in rotation or tilt, or changes on the image channel, such as brightness, treating these variables as uncertain inputs. We utilize generalized polynomial chaos (GPC) to approximate a surrogate model of the image classification model's input variables on the output [14]. The surrogate model is then employed to compute the Sobol indices, which quantify the sensitivity with respect to the perturbations and their joint interactions.

In Section 2, we recall the main ideas behind SA. In Section 3, we introduce a method to quantify the sensitivity of perturbations on the input image with respect to the output of the classification model. Section 4 validates the approach through two case studies: (i) a welding seam defect classification problem, utilizing a fine-tuned ResNet18 model and (ii) an actual image classification model for emblem detection employed in BMW Group's production facilities. We conclude with the discussion in Section 5.

2 SENSITIVITY ANALYSIS

SA aims to quantify the influence of input variables on a model's output [15, 10]. Let $\boldsymbol{y} = f(\boldsymbol{x})$ be a mathematical model with an input vector $\boldsymbol{x} = (x_1, \cdots, x_n)$ of size n and an output vector $\boldsymbol{y} = (y_1, \cdots, y_m)$ of size m. We are interested in quantifying the relative importance of how changes in each input x_i influence y_j, expressed through relative sensitivity indices [16]. Various methods have been developed to analyze a model's sensitivity. The most prominent methods are clustered into local and global SA.

Local SA methods aim to quantify the influence of individual input features on the model's output by changing specific input values while keeping other inputs constant [17]. Local SA approaches are characterized by their low computational cost and ease of application. However, they fail to fully represent nonlinear models and interactions between inputs [18]. In contrast, global SA methods aim to simultaneously assess the influence of all input variables on the model's output, taking parameter interactions into account [15]. The oldest global SA methods include regression-based techniques, wherein the regression coefficients are learned from the input variable's influence on the output. The sensitivity can be quantified using measures such as the Pearson correlation coefficient [10].

However, ML models for image classification are typically nonlinear, allowing the network to learn more complex representations of the input data. For these nonlinear models, polynomial chaos methods, such as GPC, can be used to approximate a surrogate model to the input-output data, enabling the computation of sensitivity coefficients through easier-to-interpret models [19]. The sensitivity coefficients can then be computed using variance-based methods. Variance-based methods, such as Sobol indices, decompose the variance of the output into the contributions of each input variable. This approach acknowledges that different input variables and their interactions contribute variably to the model's output variations [20].

2.1 Related work

Utilizing GPC and Sobol indices for SA has found broad adaptation in various domains such as mechanical, civil, biomedical, and chemical engineering [21, 22, 23, 24]. In the fields of material engineering and nondestructive testing, GPC has been applied to assess uncertainty in Lorentz force eddy current testing, providing valuable insights into the reliability of defect detection through Sobol indices [25]. A similar approach has been used to identify optimal stimulation sites for transcranial magnetic stimulation by evaluating sensitivities to uncertain model parameters, such as the conductivity of grey and white matter [26, 27]. For analyzing the sensitivity of image models, the authors in [28] propose using Sobol indices to quantify higher-order interactions between input features and the model's output. This approach attributes the neural network's predictions to the most important image regions, outperforming other black-box methods such as RISE (randomized input sampling for explanation) [29] and even white-box methods like Grad-CAM (gradient-weighted class activation mapping) [30].

Building on this literature, the following briefly overviews the theoretical background of GPC and Sobol indices.

2.2 Generalized polynomial chaos

GPC is a method used to approximate a surrogate model of $f(\boldsymbol{\xi})$ with random variables $\boldsymbol{\xi}$. The model is approximated by expanding it in terms of multivariate orthogonal polynomial basis functions Ψ. This can be expressed as [14, 31]

$$f(\boldsymbol{\xi}) \approx \sum_{\boldsymbol{\alpha} \in A} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}). \tag{1}$$

The expansion coefficients $c_{\boldsymbol{\alpha}}$ depend on the multi-index $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in A$ where $A \subseteq \mathbb{N}_0^d$. The multivariate orthogonal polynomial basis functions are given by $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \prod_{i=1}^d \psi_{\alpha_i}^i(\xi_i)$. These basis functions are products of univariate orthogonal polynomials $\psi_{\alpha_i}(\xi_i)$ of order α_i, which are orthogonal with respect to the probability density functions $p_{\Xi_i}(\xi_i)$ [32]. In this context, orthogonality is defined by the integral of the inner product of two basis functions, $\psi_{\alpha_i}(\xi_i)$ and $\psi_{\beta_i}(\xi_i)$, weighted by the probability measure $p_{\Xi_i}(\xi_i)$. This integral is zero unless $\alpha_i = \beta_i$. When $\alpha_i = \beta_i$, the integral equals the L^2-norm of the basis function. Therefore, the orthogonality condition can be expressed by

$$\int \psi_{\alpha_i}(\xi_i) \psi_{\beta_i}(\xi_i) p_{\Xi_i}(\xi_i) \, d\xi_i = \|\psi_{\alpha_i}(\xi_i)\|^2 \delta_{\alpha_i, \beta_i}, \tag{2}$$

where $\delta_{\alpha_i \beta_i}$ is the Kronecker delta function [19]. Using the Kronecker delta, which equals one when $\alpha_i = \beta_i$, and applying Equations (1) and (2), it can be demonstrated that the coefficients $c_{\boldsymbol{\alpha}}$ are expressed as

$$c_{\boldsymbol{\alpha}} = \frac{1}{\prod_i \|\psi_i(\xi_i)\|^2} \int f(\boldsymbol{\xi}) \Psi(\boldsymbol{\xi}) p_{\Xi}(\boldsymbol{\xi}) \, d\boldsymbol{\xi}. \tag{3}$$

The term can be approximated by regression or quadrature methods [31, 14].

By leveraging a surrogate model to approximate a model's output probability distribution using Equation (1), Sobol indices can be computed efficiently.

2.3 Sobol indices

Sobol indices are a variance-based sensitivity analysis method that quantifies the influence of a subset $\boldsymbol{\xi}_\tau$ of a model's inputs on its outputs. The indices $\tau \in \mathcal{P}(\{1, \ldots, d\})$ include all interactions of a model's input variables. Thus, Sobol indices can quantify the influence of individual input variables and the combined effect of different interactions of input variables [33]. We distinguish between total Sobol indices D_τ and relative Sobol indices S_τ. The total Sobol indices are given by

$$D_\tau = \mathbb{V}_{\boldsymbol{\xi}_\tau}[f(\boldsymbol{\xi})], \tag{4}$$

where $\mathbb{V}[f(\boldsymbol{\xi})]$ is the variance of $f(\boldsymbol{\xi})$ explained by the input variables $\boldsymbol{\xi}_\tau$. For the relative Sobol indices, we normalize Equation (4) by the sum over all interactions of input variables [15]

$$S_\tau = \frac{D_\tau}{\sum_\tau D_\tau} = \frac{\mathbb{V}_{\boldsymbol{\xi}_\tau}[f(\boldsymbol{\xi})]}{\sum_\tau \mathbb{V}_{\boldsymbol{\xi}_\tau}[f(\boldsymbol{\xi})]} = \frac{\mathbb{V}_{\boldsymbol{\xi}_\tau}[f(\boldsymbol{\xi})]}{\mathbb{V}[f(\boldsymbol{\xi})]}. \tag{5}$$

Thus, S_τ can be seen as a measure of the relative influence of $\boldsymbol{\xi}_\tau$ on $f(\boldsymbol{\xi})$. Global and relative Sobol indices can be computed using the coefficients in Equation (1) by

$$S_\tau = \frac{\sum_{\boldsymbol{\alpha} \in A_\tau} c_{\boldsymbol{\alpha}}^2 \|\Psi_{\boldsymbol{\alpha}}\|^2}{\sum_{\boldsymbol{\alpha} \in A} c_{\boldsymbol{\alpha}}^2 \|\Psi_{\boldsymbol{\alpha}}\|^2}, \tag{6}$$

with $A_\tau = \{\boldsymbol{\alpha} \in A : \alpha_i > 0 \iff i \in \tau\}$ representing the set of multi-indices $\boldsymbol{\alpha}$ for which $\alpha_i > 0$ if and only if i is an element of the subset τ [15].
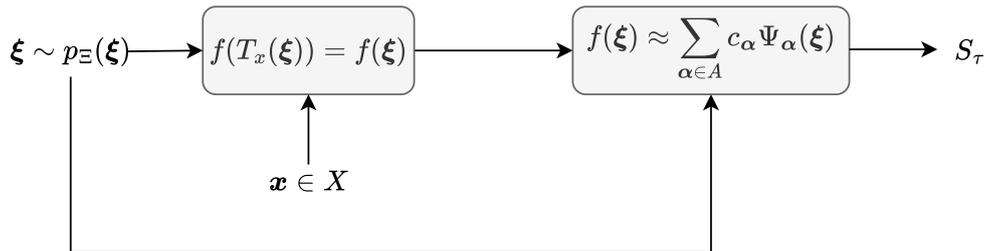
Figure 1: The fixed input $\boldsymbol{x}$, given the perturbation $\boldsymbol{\xi}$, is transformed using the transformation function $T(\boldsymbol{x}, \boldsymbol{\xi})$ and passed through the black-box model f to obtain the probability distribution $p_Y(\boldsymbol{y} \mid \boldsymbol{\xi})$. GPC approximates a surrogate model of $f(T(\boldsymbol{x}, \boldsymbol{\xi}))$ in terms of a series expansion of multivariate orthogonal polynomials. The surrogate model is then used to compute the Sobol indices S_τ.

3   GPC-SA FOR IMAGE CLASSIFICATION MODELS

Perturbations on images, such as noise, lighting variations (e.g., brightness), and geometric distortions (e.g., rotation), can significantly impact the performance of image classification models. These perturbations introduce variations in the input data distribution that the model may not be adequately trained on. As a result, the model must account for these variations to maintain robust performance.

Let $f : \mathbb{R}^{m \times n \times c} \to \{1, \dots, K\}$ define the image classification model with input images $\boldsymbol{x} \in \mathbb{R}^{m \times n \times c}$, where m and n depict the width and height of the image, and c represents the number of image channels. The set $y = \{1, \dots, K\}$ represents the output labels with an associated probability distribution $p_Y(\boldsymbol{y})$ for K classes. Now, let $p_Y(\boldsymbol{y} \mid \boldsymbol{x}) = f(\boldsymbol{\theta}, \boldsymbol{x})$ be a probabilistic classification model with learnable parameters $\boldsymbol{\theta}$ that maps $\boldsymbol{x}$ to $p_Y(\boldsymbol{y} \mid \boldsymbol{x})$ [34]. We aim to quantify the influence of perturbations on the image as transformations of the model's input variables $\boldsymbol{x}$ on its output $p_Y(\boldsymbol{y} \mid \boldsymbol{x})$ using Sobol indices of different interaction orders.

To quantify the relative influence of perturbations on the model's output, we consider a set of random variables $\boldsymbol{\xi} = (\xi_1, \dots, \xi_d)$ drawn from a probability distribution $p_\Xi(\boldsymbol{\xi})$. We define the transformation function $T(\boldsymbol{x}, \boldsymbol{\xi})$ to analyze these perturbations. We then compute the transformed inputs $\boldsymbol{x}'^{(i)} = T(\boldsymbol{x}, \boldsymbol{\xi}^{(i)})$ for each sample $\boldsymbol{\xi}^{(i)}$. The model then predicts the probability $f(\boldsymbol{x}'^{(i)}) = f(T(\boldsymbol{x}, \boldsymbol{\xi}^{(i)})) = p_Y(\boldsymbol{y} \mid \boldsymbol{x}'^{(i)})$ of $\boldsymbol{x}'$ belonging to class $\boldsymbol{y}$. By fixing $\boldsymbol{x}$, we define $T_x(\boldsymbol{\xi})$ as a function that depends only on the transformations $\boldsymbol{\xi}$ and serves as an uncertain input of the model. Thus, for a fixed $\boldsymbol{x} \in X$, it follows

$$f(\boldsymbol{x}') = f(T_x(\boldsymbol{\xi})) = f(\boldsymbol{\xi}) = p_Y(\boldsymbol{y} \mid \boldsymbol{\xi}). \tag{7}$$

This approach allows us to isolate and analyze the influence of the perturbations $\boldsymbol{\xi}$ on the model's predictions, facilitated by the computation of Sobol indices to quantify the influence of $\boldsymbol{\xi}$ on the model's output. However, the range of polynomials is $(-\infty, +\infty)$, while the range of a discrete probability distribution is $[0, 1]$. To account for this, we apply the logit function in order to map the probability to $(-\infty, +\infty)$. This can be written as

$$f(\boldsymbol{\xi}) = \text{logit}(p_Y(\boldsymbol{y} \mid \boldsymbol{\xi})) = \log\left(\frac{p_Y(\boldsymbol{y} \mid \boldsymbol{\xi})}{1 - p_Y(\boldsymbol{y} \mid \boldsymbol{\xi})} \right). \tag{8}$$

Using Equation (1), we approximate the surrogate model by

$$f(\boldsymbol{\xi}) = \text{logit}(p_Y(\boldsymbol{y} \mid \boldsymbol{\xi})) \approx \sum_{\boldsymbol{\alpha} \in A} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}). \tag{9}$$

Next, we compute the Sobol indices S_τ for various interaction orders τ for each input $\boldsymbol{x}'^{(i)}$ using Equation (6). To interpret the probabilities, we apply the inverse of Equation (8), which is the logistic function. This can be written as

$$\text{logistic}\left(\sum_{\boldsymbol{\alpha} \in A} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}\right) = \frac{1}{1 + \exp\left(-\sum_{\boldsymbol{\alpha} \in A} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}\right)}. \tag{10}$$

A schema of the complete approach is provided in Figure 1.

4 RESULTS

In the following, we validate our proposed approach using image classification models that have been fine-tuned on two different datasets. We demonstrate that approximating $f(T(\boldsymbol{x}, \boldsymbol{\xi}))$ using GPC and computing the Sobol indices can be used to quantify the sensitivity of perturbations in the input image on the model output. In Section 4.1, we analyze the sensitivity of a classification model trained on a benchmark welding seam dataset. Section 4.2 analyzes an image classification model used in BMW Group's vehicle manufacturing plants to detect the correct assembly of brand emblems.

4.1 TIG welding seam classification

We validate our approach on an image classification model using a pre-trained and fine-tuned ResNet18 model [35]. The model comprises 11.7 million parameters, ReLU activations, and a single 7x7 convolution layer with pooling. The pre-trained ResNet18 model is fine-tuned on benchmark Aluminum 5083 TIG welding seam images recorded with an HDR camera, down-sampled to 224x224 pixels with a single channel and without applying any transformations [36]. The training dataset contains 26,666 images, and the test dataset contains 6,588 images. The images show six defect types: good weld, burn-through, contamination, lack of fusion, misalignment, and lack of penetration.

Table 1: Sobol indices from the GPC surrogate model for *brightness*, *rotation*, and *tilt*. The results for all three transformations have a relative influence of 0.061. *Tilt* exhibits the highest influence.

	brightness	rotation	tilt
brightness	0.171	0.03	0.162
rotation	-	0.11	0.096
tilt	-	-	**0.37**

We are interested in evaluating the sensitivity of the fine-tuned image classification model to variations in brightness, rotation, and tilt, as well as the interactions among these transformations. To achieve this, we compute the Sobol indices, which quantify the contribution of each input variable and their interactions with the output variance. By analyzing the Sobol indices, we can determine the sensitivity of the classification model to each transformation and understand how these transformations interact to affect the model's performance.

For this purpose, we draw $n = 1000$ samples using Latin hypercube sampling (LHS) from a Beta distribution with specified limits for each transformation parameter. The

(A) Surface plot GPC approximation   **(B)** GPC approximation brightness
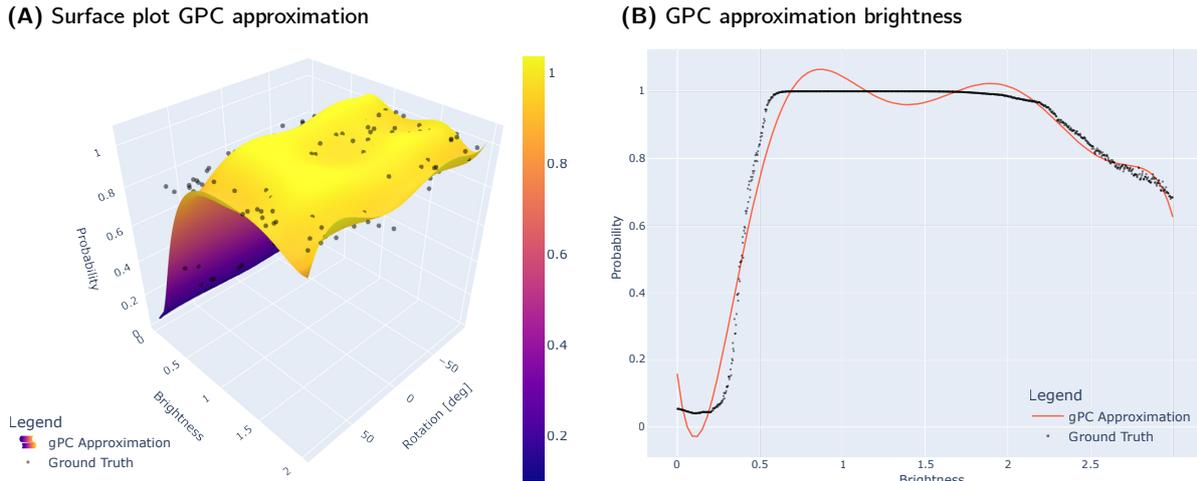

Figure 2: Figure (A) presents a surface plot of the GPC approximation, illustrating the relationship between *brightness* and *rotation* for the prediction probability, with *tilt* held constant at 0. Figure (B) displays the approximation of *brightness* while both *rotation* and *tilt* are fixed at 0.

limits are $[0, 2]$ for *brightness* (representing the scaling factor of brightness change, where 1 represents the original brightness), $[-30, 30]$ for *rotation* (the angle of rotation in degrees), and $[-20, 20]$ for *tilt* (the angle of tilt in degrees).

We apply these transformations to a fixed input image and compute the output probabilities of the image classification model using Equation (7). The surrogate model is approximated using GPC with Jacobi polynomials. The relative error of the approximation is 0.057. We set the maximum total order of polynomials to four and use the Moore-Penrose pseudo inverse to compute the polynomial coefficients. We then determined the Sobol indices with the computed coefficients, as given by Equation (6).

Table 1 presents the resulting Sobol indices. These indices indicate that *tilt* has the highest influence on the model, with a value of 0.37. In contrast, *brightness* and *rotation* exert the lowest influence on the model, each with a value of 0.03. The combined relative influence of *brightness*, *rotation*, and *tilt* is 0.061[1].

The surrogate models illustrating the GPC approximation for *rotation* and *brightness* are depicted in Figure 2. Figure 2 (A) presents a surface plot of the surrogate model with respect to *rotation* and *brightness*, where *tilt* is fixed at a value of 0. Figure 2 (B) shows a plot of the surrogate model with respect to *brightness*, where *rotation* and *tilt* are fixed at a value of 0.

4.2   Vehicle emblem classification

In this experiment, we validate our SA approach on an image classification model fine-tuned to detect the correct assembly of brand emblems in BMW's Group vehicle production plants. The model classifies images into three distinct emblem types and includes an additional class for unclassified instances. Operating as a black box, the image classification model provides output class probabilities for given inputs. We aim

---

[1]For those interested in reproducing the results or exploring the SA approach further, the code including the fine-tuned ResNet18 model is available on GitHub at `https://github.com/lpossner/GPC_Image_Classification`.

(A) Surface plot GPC approximation
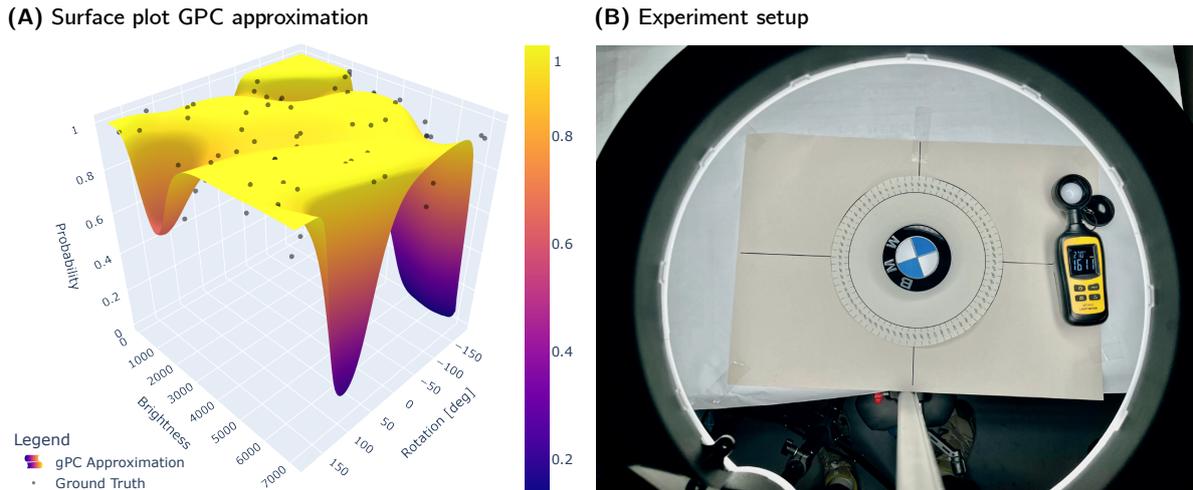
(B) Experiment setup

Figure 3: Figure (A) presents a surface plot of the GPC approximation, illustrating the relationship between *brightness* and *rotation* for the prediction probability. Figure (B) shows the experimental setup used to obtain the validation images.

to evaluate its sensitivity to variations in *rotation* and *brightness* and the interactions between these transformations, using Sobol indices.

To conduct our experiment, we set up an apparatus comprising a Keyence camera, an adjustable light source, a luminance meter, and an angle scale. We then adjusted the *rotation* and *brightness* for $n = 100$ samples using LHS drawn from a Beta distribution. The limits for the *rotation* transformation (the angle of rotation in degrees) are set between $[-180, 180]$, and for the *brightness* transformation (representing brightness in Lux), the range is

Table 2: Result of the Sobol indices for *brightness* and *rotation*. The joint interaction exhibits the highest influence on the emblem classification model.

	brightness	rotation
brightness	0.229	**0.403**
rotation	-	0.368

$[0, 7000]$. With this configuration, we adjust the *rotation* angle and *brightness* and shot the images (cf. Figure 3 (B)) to compute the output class probabilities using Equation (7). To approximate the surrogate model, we employ GPC with Jacobian polynomials and subsequently compute the Sobol indices. The relative error obtained from this method is 0.23, given a maximum total polynomial order of $[6, 5]$.

Table 2 presents the Sobol indices obtained from our experiment. The results indicate that the joint interaction of *brightness* and *rotation* has the highest influence, with a Sobol index of 0.403. Figure 3 (A) displays a surface plot of the GPC approximation. In the following chapter, we contextualize the results.

5 DISCUSSION AND CONCLUSION

This paper presents preliminary results on the effectiveness of sensitivity analysis of image classification models with perturbed images using GPC. We introduce a method where a fixed input image is transformed by samples from a known distribution that is associated with a family of orthogonal polynomial functions. Using the output of the image classification model for the transformed input, we employ GPC to approximate a

surrogate model [14]. This surrogate model enables us to determine the relative influence of each transformation applied to the input on the model's output, quantified as Sobol indices.

In Section 4, we demonstrate the approach on an Aluminum 5083 TIG welding seam images dataset for which we fine-tuned a ResNet18 model and on an emblem image classification model utilized in BMW Group's production facilities [36]. The experiment provides initial results on evaluating the models' sensitivity to images that are exposed to perturbations, which can potentially lower the model's prediction performance. Our findings show we can approximate the nonlinear models' input-output relationships using GPC. The findings further suggest that geometric transformations and brightness adjustments alter the visual structure in a way that negatively affects the prediction accuracy of the classification model. This can be attributed to distributional domain shifts that differ from the models' training dataset. Quantifying the relative influence of input transformations using Sobol indices provides valuable insights into how variations in the input affect the model's predictions. This information can assist practitioners in enhancing the robustness of models, for example, making the emblem classification model more resilient to different lighting conditions.

However, several questions remain unanswered. It remains to be validated whether the Sobol indices generalize across the entire distribution of the dataset space for a fixed input, suggesting a potential avenue for future research. Moreover, in this study, we test only one family of polynomials, specifically the Jacobian polynomials. This leaves potential research opportunities open to exploring other polynomial families and examining their impact on the accuracy and efficiency of the surrogate model. Further investigation is required to determine the robustness of our method across different datasets and GPC configurations.

REFERENCES

[1] ISO/IEC, "Risk management — Risk assessment techniques," standard, International Organization for Standardization, Geneva, CH, June 2019.

[2] IEC, "IEC 60812:2018 - Failure modes and effects analysis (FMEA and FMECA)," tech. Rep., International Electrotechnical Commission (IEC), Geneva, Switzerland, 2018.

[3] E. Ruijters and M. Stoelinga, "Fault tree analysis: A survey of the state-of-the-art in modeling, analysis and tools," *Computer Science Review*, vol. 15-16, pp. 29–62, Feb. 2015.

[4] L. Bahr, C. Wehner, J. Wewerka, J. Bittencourt, U. Schmid, and R. Daub, "Knowledge Graph Enhanced Retrieval-Augmented Generation for Failure Mode and Effects Analysis," *arXiv preprint arXiv:2406.18114*, July 2024.

[5] C. A. Escobar, M. E. McGovern, and R. Morales-Menendez, "Quality 4.0: a review of big data challenges in manufacturing," *Journal of Intelligent Manufacturing*, vol. 32, pp. 2319–2334, Dec. 2021.

[6] A. Mayr, D. Kißkalt, M. Meiners, B. Lutz, F. Schäfer, R. Seidel, A. Selmaier, J. Fuchs, M. Metzner, A. Blank, and J. Franke, "Machine Learning in Production – Potentials,

Challenges and Exemplary Applications," *Procedia CIRP*, vol. 86, pp. 49–54, Jan. 2019.

[7] A. Zonnenshain and R. S. Kenett, "Quality 4.0—the challenging future of quality engineering," *Quality Engineering*, vol. 32, pp. 614–626, Oct. 2020.

[8] L. Possner, L. Bahr, L. Roehl, C. Wehner, and S. Groeger, "Anwendung von Causal-Discovery-Algorithmen zur Root-Cause-Analyse in der Fahrzeugmontage," *arXiv preprint arXiv:2407.16388*, July 2024.

[9] H. Tercan and T. Meisen, "Machine learning and deep learning based predictive quality in manufacturing: a systematic review," *Journal of Intelligent Manufacturing*, vol. 33, pp. 1879–1905, Oct. 2022.

[10] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models.* Wiley, 1 ed., Feb. 2002.

[11] V. T. Vasudevan, A. Sethy, and A. R. Ghias, "Towards Better Confidence Estimation for Neural Models," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7335–7339, 2019.

[12] Y. Ovadia, E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J. V. Dillon, B. Lakshminarayanan, and J. Snoek, "Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Red Hook, NY, USA: Curran Associates Inc., 2019.

[13] J. Gawlikowski, C. R. N. Tassi, M. Ali, J. Lee, M. Humt, J. Feng, A. Kruspe, R. Triebel, P. Jung, R. Roscher, M. Shahzad, W. Yang, R. Bamler, and X. X. Zhu, "A survey of uncertainty in deep neural networks," *Artificial Intelligence Review*, vol. 56, pp. 1513–1589, Oct. 2023.

[14] K. Weise, L. Poßner, E. Müller, R. Gast, and T. R. Knösche, "Pygpc: A sensitivity and uncertainty analysis toolbox for Python," *SoftwareX*, vol. 11, p. 100450, Jan. 2020.

[15] B. Sudret, "Global sensitivity analysis using polynomial chaos expansions," *Reliability Engineering & System Safety*, vol. 93, pp. 964–979, July 2008.

[16] E. Borgonovo and E. Plischke, "Sensitivity analysis: A review of recent advances," *European Journal of Operational Research*, vol. 248, pp. 869–887, Feb. 2016.

[17] A. Saltelli and P. Annoni, "How to avoid a perfunctory sensitivity analysis," *Environmental Modelling & Software*, vol. 25, pp. 1508–1517, Dec. 2010.

[18] Y. Tang, P. Reed, T. Wagener, and K. van Werkhoven, "Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation," *Hydrology and Earth System Sciences*, vol. 11, pp. 793–817, Feb. 2007.

[19] D. Xiu and G. Em Karniadakis, "Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos," *Computer Methods in Applied Mechanics and Engineering*, vol. 191, pp. 4927–4948, Sept. 2002.

[20] A. Saltelli, "Sensitivity analysis for importance assessment.," *Risk analysis : an official publication of the Society for Risk Analysis*, vol. 22, pp. 579–590, June 2002.

[21] D. Lucor, J. Meyers, and P. Sagaut, "Sensitivity analysis of large-eddy simulations to subgrid-scale-model parametric uncertainty using polynomial chaos," *Journal of Fluid Mechanics*, vol. 585, pp. 255–279, Aug. 2007.

[22] P. Ni, Y. Xia, J. Li, and H. Hao, "Using polynomial chaos expansion for uncertainty and sensitivity analysis of bridge structures," *Mechanical Systems and Signal Processing*, vol. 119, pp. 293–311, Mar. 2019.

[23] P. L. T. Duong, W. Ali, E. Kwok, and M. Lee, "Uncertainty quantification and global sensitivity analysis of complex chemical process using a generalized polynomial chaos approach," *Computers & Chemical Engineering*, vol. 90, pp. 23–30, July 2016.

[24] M. Tosin, A. M. A. Côrtes, and A. Cunha, "A Tutorial on Sobol' Global Sensitivity Analysis Applied to Biological Models," in *Networks in Systems Biology: Applications for Disease Modeling*, pp. 93–118, Cham: Springer International Publishing, 2020.

[25] K. Weise, M. Carlstedt, M. Ziolkowski, and H. Brauer, "Uncertainty Analysis in Lorentz Force Eddy Current Testing," *IEEE Transactions on Magnetics*, vol. 52, pp. 1–4, Mar. 2016.

[26] G. B. Saturnino, A. Thielscher, K. H. Madsen, T. R. Knösche, and K. Weise, "A principled approach to conductivity uncertainty analysis in electric field calculations," *NeuroImage*, vol. 188, pp. 821–834, Mar. 2019.

[27] O. Numssen, K. Weise, A. Thielscher, T. R. Knösche, and G. Hartwigsen, "Determining the stimulation site in TMS: Sensitivities with respect to uncertain model parameters," June 2019.

[28] T. Fel, R. Cadène, M. Chalvidal, M. Cord, D. Vigouroux, and T. Serre, "Look at the variance! efficient black-box explanations with sobol-based sensitivity analysis," in *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, (Red Hook, NY, USA), Curran Associates Inc., 2024.

[29] V. Petsiuk, A. Das, and K. Saenko, "RISE: Randomized Input Sampling for Explanation of Black-box Models," *arXiv:1806.07421*, 2018.

[30] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," *International Journal of Computer Vision*, vol. 128, pp. 336–359, Feb. 2020.

[31] D. Xiu and G. E. Karniadakis, "The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations," *SIAM Journal on Scientific Computing*, vol. 24, no. 2, pp. 619–644, 2002.

[32] R. Askey and J. A. Wilson, *Some basic hypergeometric orthogonal polynomials that generalize Jacobi polynomials*, vol. 319. American Mathematical Soc., 1985.

[33] I. M. Sobol, "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates," *Mathematics and Computers in Simulation*, vol. 55, pp. 271–280, Feb. 2001.

[34] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning.* MIT Press, 2016.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[36] D. Bacioiu, G. Melton, M. Papaelias, and R. Shaw, "Automated defect classification of Aluminium 5083 TIG welding using HDR camera and neural networks," *Journal of Manufacturing Processes*, vol. 45, pp. 603–613, Sept. 2019.