UNSUPERVISED LEARNING FOR HIGH-FIDELITY COMPRESSION OF LARGE EXPERIMENTAL DATASET: AN APPLICATION ON HPT BLADE TIP CONTOURING

F. A. Tucci[1] - G. Delibra[1,] - L. Tieghi[1] - A. Corsini[1] - S. Lavagnoli[2]*

[1]Department of Mechanical and Aerospace Engineering, Sapienza University, Rome, Italy,
[2]Turbomachinery and Propulsion Department, von Karman Institute for Fluid Dynamics, Sint-Genesius Rode, Belgium
[*]giovanni.delibra@uniroma1.it

ABSTRACT

Rotor tip contouring is one of the most complex and promising methods to enhance the performance of high-pressure turbines (HPT). In this study, a new unsupervised machine learning approach based on dimensionality reduction is presented. Derived from methods developed for image recognition technology, this approach explores the functional relationships between rotor tip contouring and aero-thermodynamic performance of HPT rotors. The procedure is applied to experimental measurements carried out on the rotating turbine rig at von Karman Institute, which operates a rotor with multiple tip geometries divided over seven sectors.

KEYWORDS

Gas turbine, tip contouring, unsupervised learning, auto encoders, principal component analysis

NOMENCLATURE

AE	Autoencoders	ML	Machine Learning
AI	Artificial Intelligence	MSE	Mean Squared Error
ANN	Artificial Neural Network	MLP	Multi-layer Perceptron
BD	Big-Data	PCA	Principal Components Analysis
CPS	Cyber-Physical Systems	SSIM	Structural Similarity Index
HPT	High Pressure Turbine	VAE	Variational Autoencoder

INTRODUCTION

Industry 4.0 refers to a new industrial paradigm based on Big Data (BD) and Artificial Intelligence (AI) for the creation of cyber-physical systems (CPS) (Akyurt et al., 2020). Through vertical, horizontal, and end-to-end integration of industrial processes, CPS aims at establishing a dynamic environment where smart technologies can exchange data, assess production issues, and boost productivity while reducing economic, time, and energy costs (Raj et al., 2020).

In the past decade, data from engineering processes have grown in scale, since CPS-related technologies are primarily based on sharing and analyzing data from an entire industrial production process (Singh, 2021). The exploration of these unorganized data pools has demanded the development of BD tools, to explore and reorder data collected from massive and heterogeneous resources (Wu et al., 2013). With respect to traditional datasets, BD involve both structured and unstructured data and require more time and complex resources for in-depth analysis (Russom et al., 2011). BD analytics are now strategically found in most engineering applications, including environmental monitoring (Alic et al., 2019), smart grids for renewable energy networks (Quintero et al., 2021), manufacturing process optimization (Majeed et al., 2021), turbulence research (Tieghi et al., 2021), identification and quantification of loss mechanisms in turbomachinery design (Corsini et al., 2021), condition-based maintenance and predictive maintenance based on SCADA signals (Miele et al., 2022), health care applications (Karatas et al., 2018) and face recognition (Yang and Song, 2022).

The turbomachinery community is also reconsidering manufacturing in light of the exponential growth in the use of smart technologies, particularly those connected to the creation of more complex solutions to boost core engine efficiency (Xie et al., 2021). To this end, the use of surrogate models to speed up prediction time and lower the overall number of numerical simulations and experiments needed to perform design optimization is the state-of-the-art in turbomachinery design at the time of writing (Boursier et al., 2018). In fact, standard approaches may fail in capturing the complexity of the underlying physics, as the modelling accuracy is limited by the number of features. Engineers frequently rely on empirical approaches and review of analysis graphs to develop design strategies because it is challenging or impracticable to explicitly describe these features in advance. Using AI models that can consider every feature required for design optimization may constitute an answer. In fact, in recent years, the adoption of more sophisticated sensors and the growth in computing power have made both high-resolution numerical simulations and experimental measurements possible, making a vast quantity of data on turbomachinery performance available. As an example, Cernat et al. (2018) performed experiments to measure the unsteady aerothermal field of various blade tip geometries, reporting measurements of static pressure and heat transfer on the turbine rotor shroud that were both time-averaged and time-resolved. Biedermann et al. (2019) carried out another extensive experimental work to understand how sinusoidal leading edges of rotors modify the acoustic signature of industrial fans. Zhang et al. (2015) experimentally investigated the trajectory and dynamics of the tip leakage vortex in axial flow pumps. Lavagnoli et al. (2012) describe the full implementation of a high-frequency capacitive sensor on the shroud of a large transonic turbine stage.

As a result, today's challenge is to assess the capabilities and limits of BD approaches in turbomachinery performance analysis, with the goals of unveiling useful information and assist the design process. However, in the exploration of large datasets it is essential to reduce the number of features to a selection that retains most of the original information. This process is called dimensionality reduction and prevents dealing with a large amount of data and complicating AI models. Principal Components Analysis (PCA) and Variational Autoencoders (VAEs), a probabilistic form of Autoencoders, are among the most used approaches for BD dimensionality reduction. In his research, Zagler (2017) trained a VAE to simulate the characteristics of compressor blade supersonic airflow. Pongetti et al. (2021) proved how AEs can be used to provide compressed form insights from the complete flow field in forecasts of quantities of interest, to improve the accuracy of AI models employed in high load transonic compressor blade optimization cycles. Angelini et al. (2020) developed a dimensionality-reduced parametrization of complex tip geometries based on PCA scores for gas turbine rotors, overcoming the limitations of discrete topology approaches.

In similar multi-scale and multi-physic problems, it is difficult to find analytical models to correlate the geometric parametrization of the blade tip and the aerodynamic and thermal performance. Therefore, this paper proposes a Machine Learning (ML) methodology based on a combination of Autoencoders and PCA applied to large experimental data sets to investigate the relationships between tip geometry and unsteady aero-thermodynamic performance in HPT blading.

The approach is validated on experimental data measured at the von Karman Institute's rotating turbine rig, which mounts a *rainbow* rotor. Specifically, the rainbow rotor features 48 blades with the same baseline profile and distributed in 7 distinct circumferential sectors based on the similarity of their different tip geometries. In fact, each of the sectors hosts six or seven blades with similar tip geometry among them, including: single squealer, multi-cavity squealer or contoured profile. This testing method allows simultaneous measurements on different tip geometries.

METHODOLOGY

This manuscript reports on the use of an unsupervised ML workflow that, advocating AI methods developed in the field of image recognition, aims at deriving functional relationships between tip contouring geometry and aero-thermodynamic performance of HPT blades. The unsupervised ML workflow is summarized in Figure 1. The method:

a) exploits an experimental dataset of phase-locked averaged (PLA) heat transfer and pressure measurements of HPT rotors with different tip contouring;
b) entails over-sampling of both geometry and measurements;
c) provides unsupervised dimensionality reduction by means of PCA on the tip geometries, and AEs on the performance;
d) trains an artificial neural network (ANN), to predict the latent representation of performance data derived from the AE using the Principal Components of the tip geometries as input;
e) finally decodes the latent representation and compute the corresponding performance maps.

The accuracy of the results is evaluated against the original data using Structural-Similarity Index Measure (SSIM) as performance index (Wang et al., 2004).

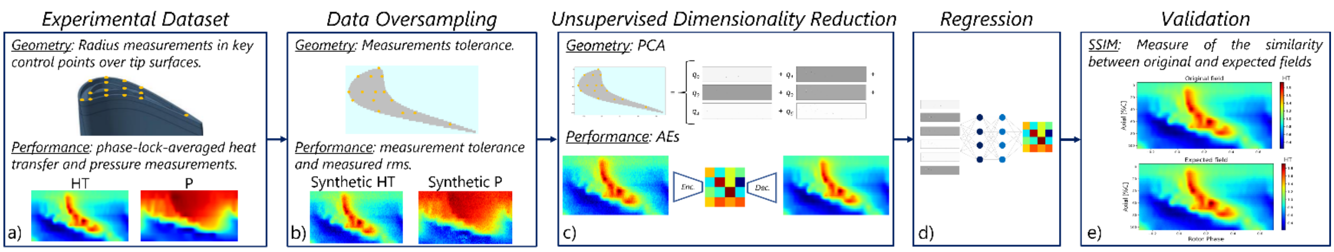
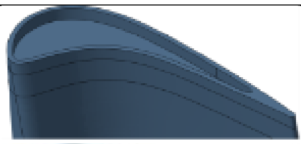


Figure 1 – Unsupervised Machine Learning Method workflow.

Experimental dataset

Table 1 provides details of the tip designs that characterize the non-confidential sector geometries. Available data include both information on the reduced tip geometries and aero-thermodynamic measurements. The static pressure field above the tip is acquired with an instrumented insert equipped with 66 pressure taps on the casing: 33 fast-response piezo-resistive pressure sensors and 33 pneumatic lines sampled by an external pressure scanner. The unsteady casing heat flux is measured by 35 double layer thin film gauges (Cernat et al., 2018).

Table 1 - Details of tip designs.

Tip Geometry Type	Tip Gap [%]	Sector Denomination	Baseline
Squealer	1.38	Sector A	
Multi-cavity squealer type	1.38	Sector B	
Multi-cavity squealer type	1.38	Sector C	

Data are mapped on a measurement plane and in the following will be treated as an image. Reduced tip geometry, Figure 2, refers to radial measurements at significant control locations on each of the 48 blades. Depending on the baseline profile, the control points on the tip surface can range from 10 to 16. The aero-thermodynamic performance is investigated using phase-lock-averaged (PLA) pressure and heat transfer measurements in each of the 7 sectors (Figure 3).

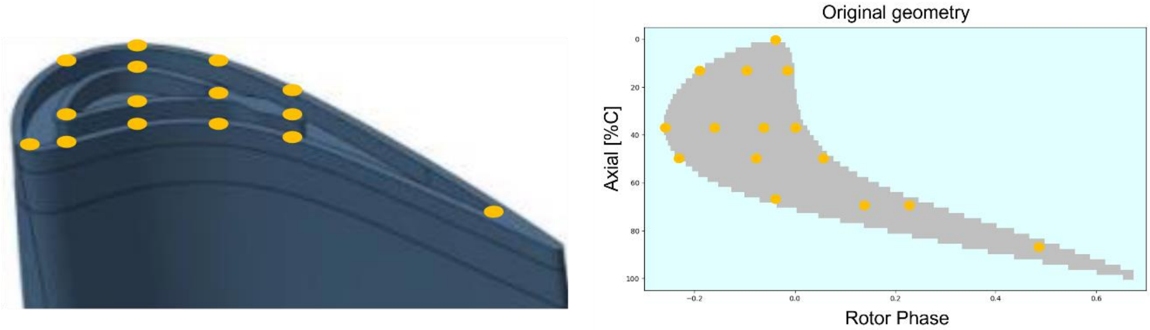


Figure 2 - Key control points for the reduced tip geometry of the B-sector baseline profile.

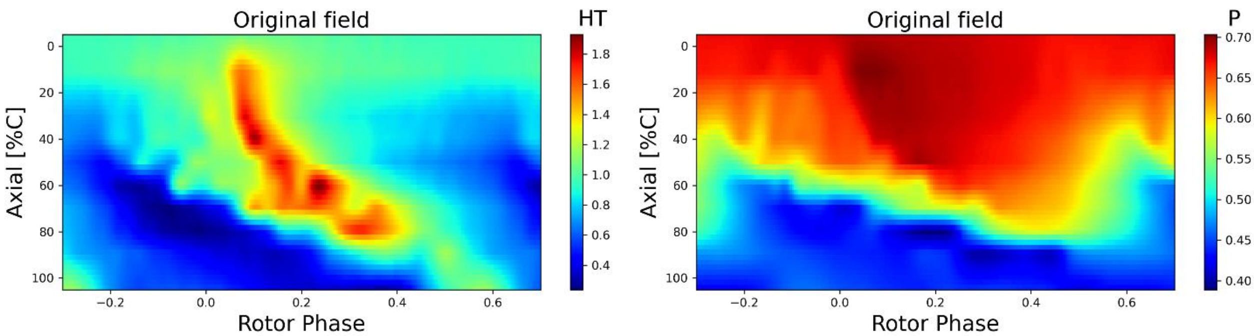


Figure 3 – Phase-locked averaged measurements of heat transfer and pressure for sector C.

Data oversampling

The automatic detection and characterization of the blade tip shape within the geometry's point-cloud is the first step in the unsupervised machine learning process. This is accomplished with a Python function that identifies the tip perimeter and assesses whether a grid point fits within it. The routine is based on the Concave Hull algorithm (Moreira et al., 2007). The algorithm automatically assigns the points that are identified in this way either the corresponding radius (59 mm in the test case) or a radius of 0 mm if it lies outside the tip perimeter. Furthermore, if a point is coincident with one of the key points in Figure 2, it assigns the real radius measurement. In the second phase of the framework, synthetic data for both performance and geometry are generated to enrich the training dataset. Traditional ML approaches enrich training dataset with different operation on images, such as cropping, rotating and resizing. However, this approach is not feasible for fluid dynamic fields in a fixed reference system. Therefore, the dataset was oversampled using a gaussian distribution of both geometry and performance. Gaussian distributions are based on the measured rms for pressure and the tolerances of measuring instruments for tip representation and heat transfer. Tip measurements have an uncertainty of ±0.005 mm, while the heat flux measurement have a tolerance of 6%, mainly due to uncertainties in thermal properties. As summarized in Table 2 the synthetic radius (R) is the sum of the actual radius ($radius_{actual}$) and its randomized tolerance (r_1). The same applies to the synthetic heat transfer (HT), which is the sum of the real heat transfer (ht_{actual}) and its randomized tolerance (ht_1), while the synthetic pressure (P) is the sum of the static pressure (p_{static}) and the randomized rms_p (p_1). The generated synthetic database entails 5952 combinations of geometry and performance: 124 combinations for each original blade. Figure 4 provides a comparison between synthetic heat transfer and pressure data with the original data.

Table 2 - Synthetic data generation criterion.

	Radius	Heat Transfer	Pressure
Random Data Generation	$r_1 = N \times 0.005,\ N\in(-1,1)$	$ht_1 = N \times 0.06,\ N\in(-1,1)$	$p_1 = rms_p \times N,\ N\in(-1,1)$
Synthetic Data	$R = radius_{actual} + r_1$	$HT = ht_{actual} + ht_1$	$P = p_{static} + p_1$

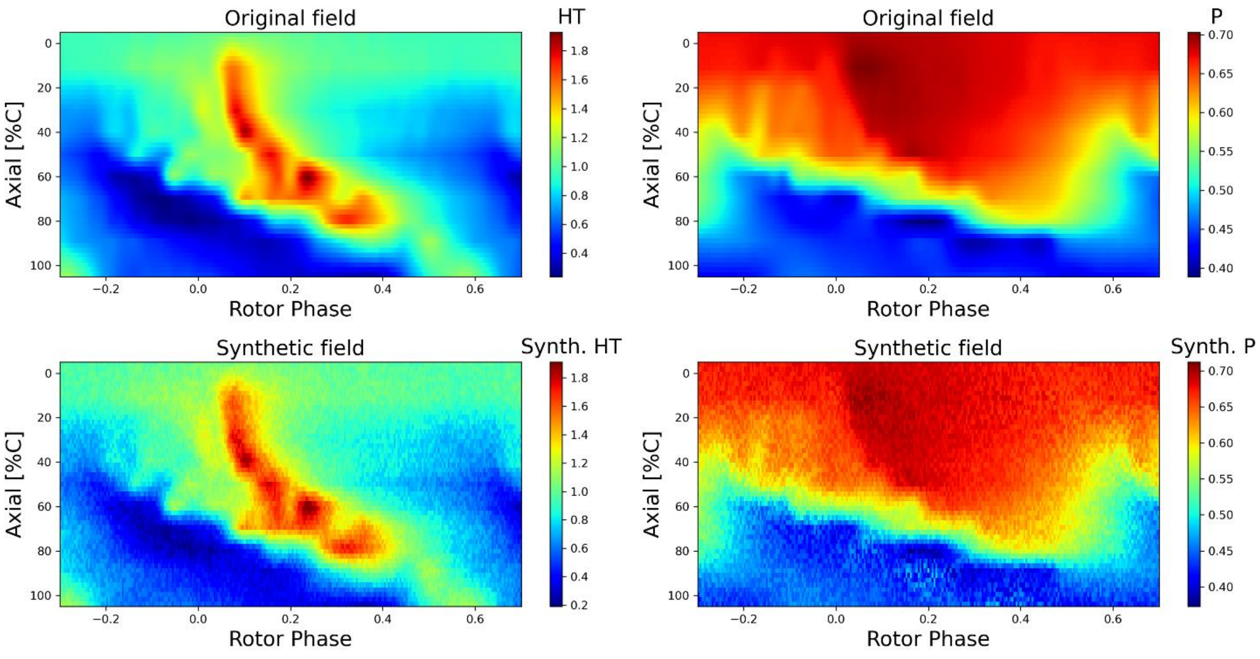


Figure 4 - Synthetic and original performance fields comparison of 1 out of 124 pairs of combinations for the 1st blade in sector C.

Unsupervised Dimensionality Reduction

Principal Components Analysis

The synthetic database entails 5952 images of possible tip characterizations each with 11781 features, i.e., image pixels. The dataset is then processed with principal component analysis (PCA) (Zhao et al., 2006). This decomposition derives a dimensionally reduced latent representation of each geometry. This compression is achieved by projecting the data in a new coordinate system in which the basis vectors are the directions with maximum variance. The number of bases is chosen to limit the loss of information. Images represented as a matrix of pixels are reshaped as a 1D vector. The total number of pixels determines the dimension of the vector space.

Six principal components are required for the training dataset to reduce the information loss below 1%, as shown by the elbow diagram in Figure 5. The corresponding loadings, also known as eigen-images (eigen-blades) are shown in Figure 6. Now each synthetic tip geometry in the database can be obtained as a linear combination of the original eigen-blades, weighted by their own score.

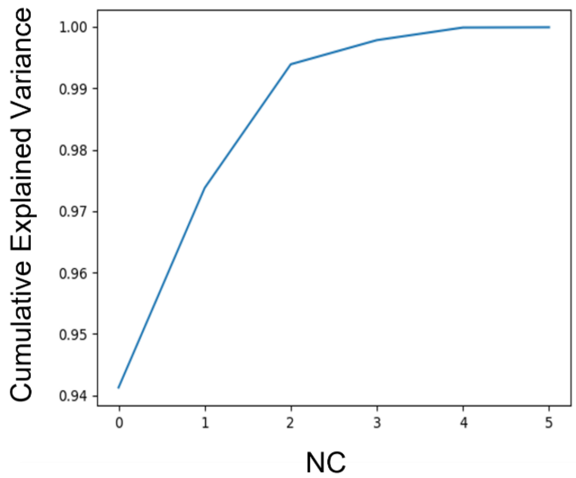


Figure 5 - PCA elbow chart.

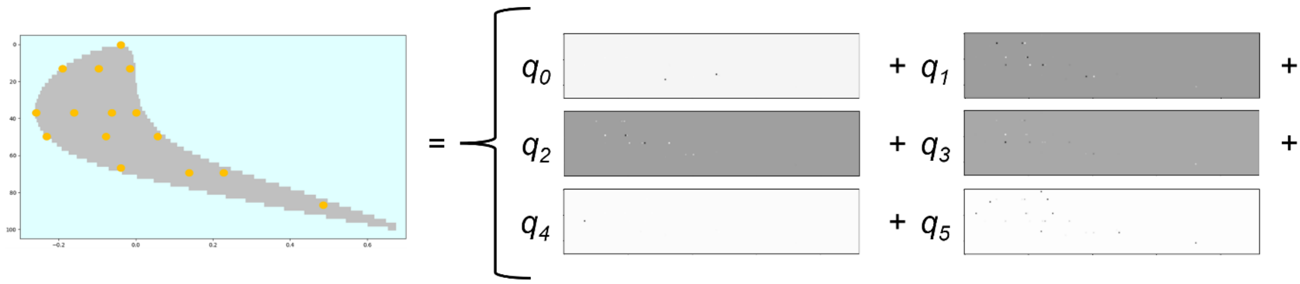


Figure 6 - PCA graphical illustration.

Autoencoders

Autoencoders (AEs) are a special class of neural networks trained to learn a compressed representation of data in an unsupervised manner, to copy its input to its output from a reduced encoding latent representation (Baldi, 2012). AEs consist of two parts: an encoder function $h=f(x)$ trained to map the input x to its latent representation h, and a decoder that produces a reconstruction function $r=g(h)$, that maps back the latent representation back to the original form. The compression-reconstruction process generates an error between reconstructed and original data. The minimization of this error is the objective function of AE training.

In the reference dataset, mean squared error (MSE) is chosen as reconstruction error metric. The encoding process is capable of derive a latent representation of HT and pressure with a dimensionality of 16 features. The decoder then maps back from the latent space to the original data. Figure 7 shows two examples of dimensionality reduction results: the upper part of the image for the heat transfer field of 1 of the 124 representations of the 1st blade of sector C, while the lower part shows the same procedure for 1 synthetic pressure field out 124 of the 4th blade of sector B.

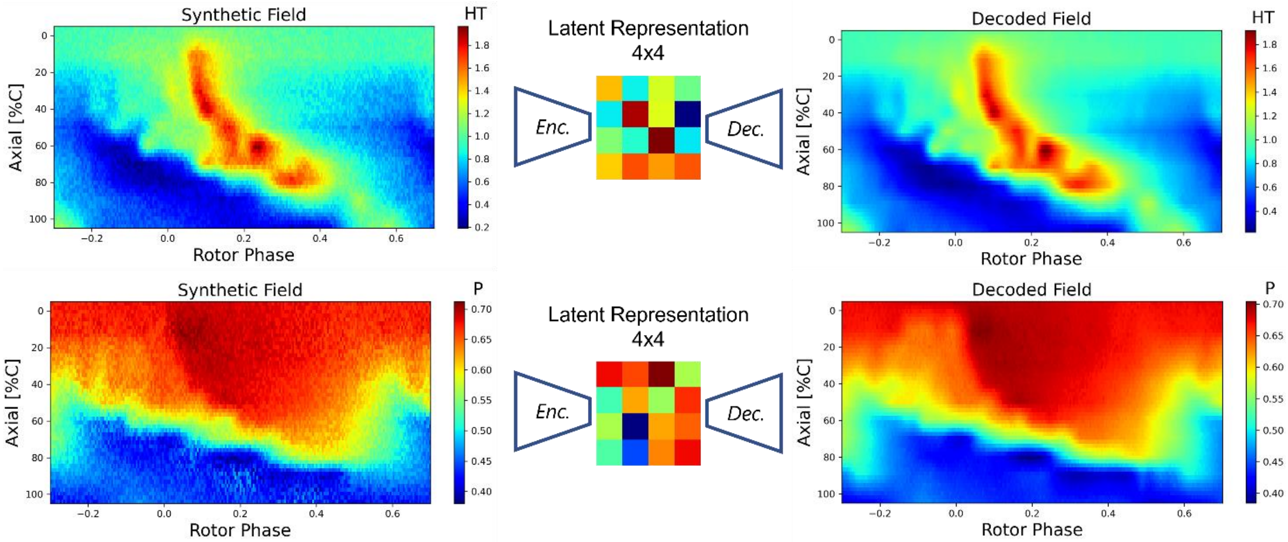


Figure 7 - Encoding and decoding procedure through AEs.

Regression

The compressed data can be more easily exploited for turbomachinery design and optimization processes. For example, in this reference dataset it is possible to correlate the blade tip contouring with the flow fields. This is done in a latent space, treating geometry PCA scores and the encoded features from AEs as input/output features. The relationship is here build using an ANN. Table 3 details the final setup of the Multi Layer Perceptron (MLP) network model optimized through a grid search algorithm (Bergstra et al., 2012), considering early stopping to avoid overfitting.

Table 3 - Hyperparameters of the MLP

Number hidden of layers	2
Number of Neurons	32
Number of training epochs	1000
Activation function	relu
Initial learning rate	1E-3
Optimizer	ADAM
Cost function	MSE
Batch size	250

Figure 8 shows a schematic view of a uses of the compressed data, with the creation of a functional relationship between tip-geometry and pressure/HT fields.

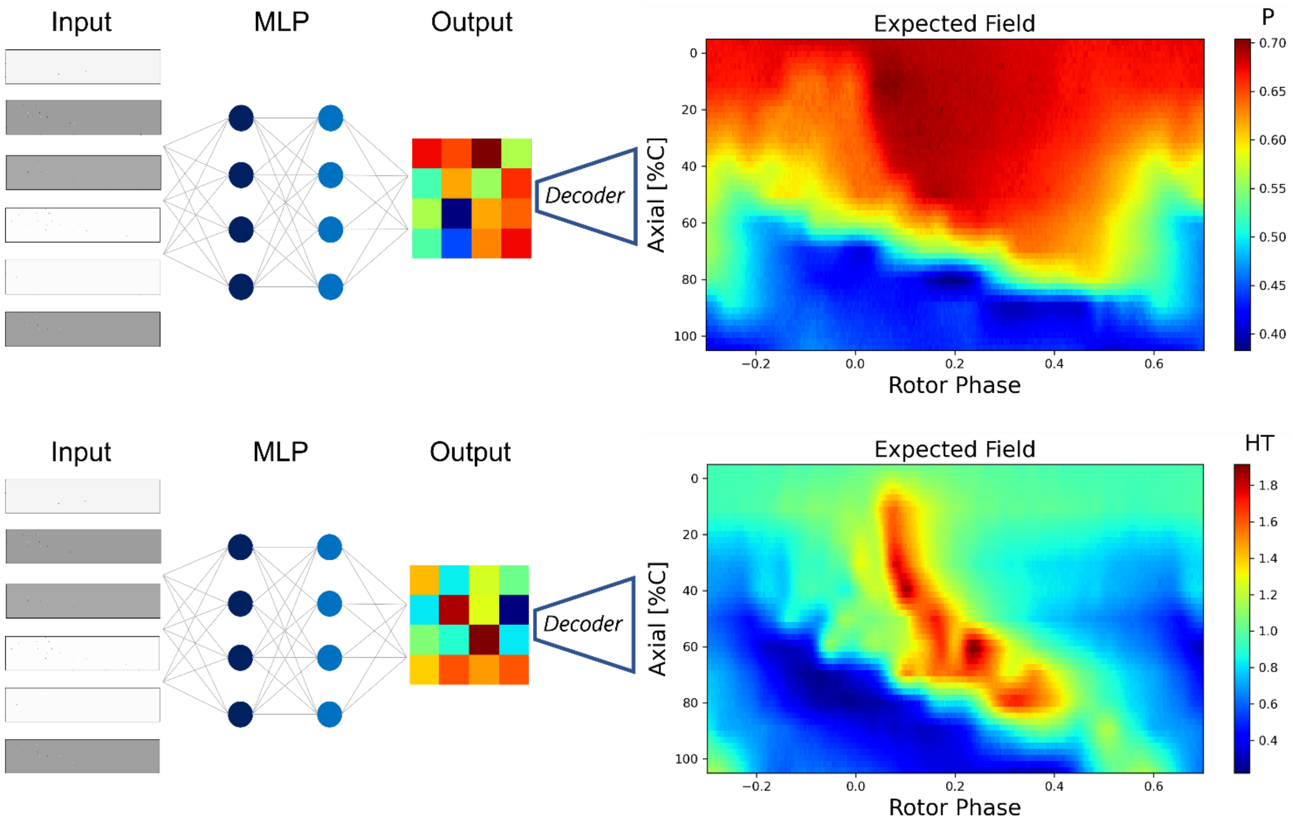


Figure 8 - Example of network for predicting expected pressure (top) and heat transfer fields (bottom).

This process can be also reversed to discover, given a desired field, the geometry that is more likely to provide this distribution, Figure 9.

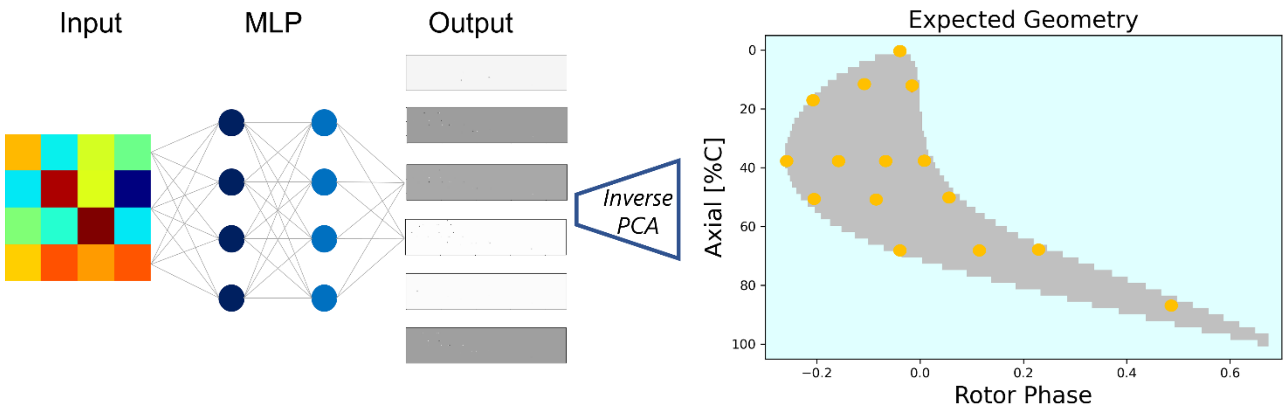


Figure 9 - Example of network for predicting the expected reduced Rainbow rotor tip geometry.

Validation

Results obtained from the validation of the framework are shown in Figure 10. The agreement between reconstructed and experimentally sampled data has been evaluated by means of the Structural Similarity Index Measure (SSIM). SSIM is used to assess the similarity between two images (Wang et al., 2004).

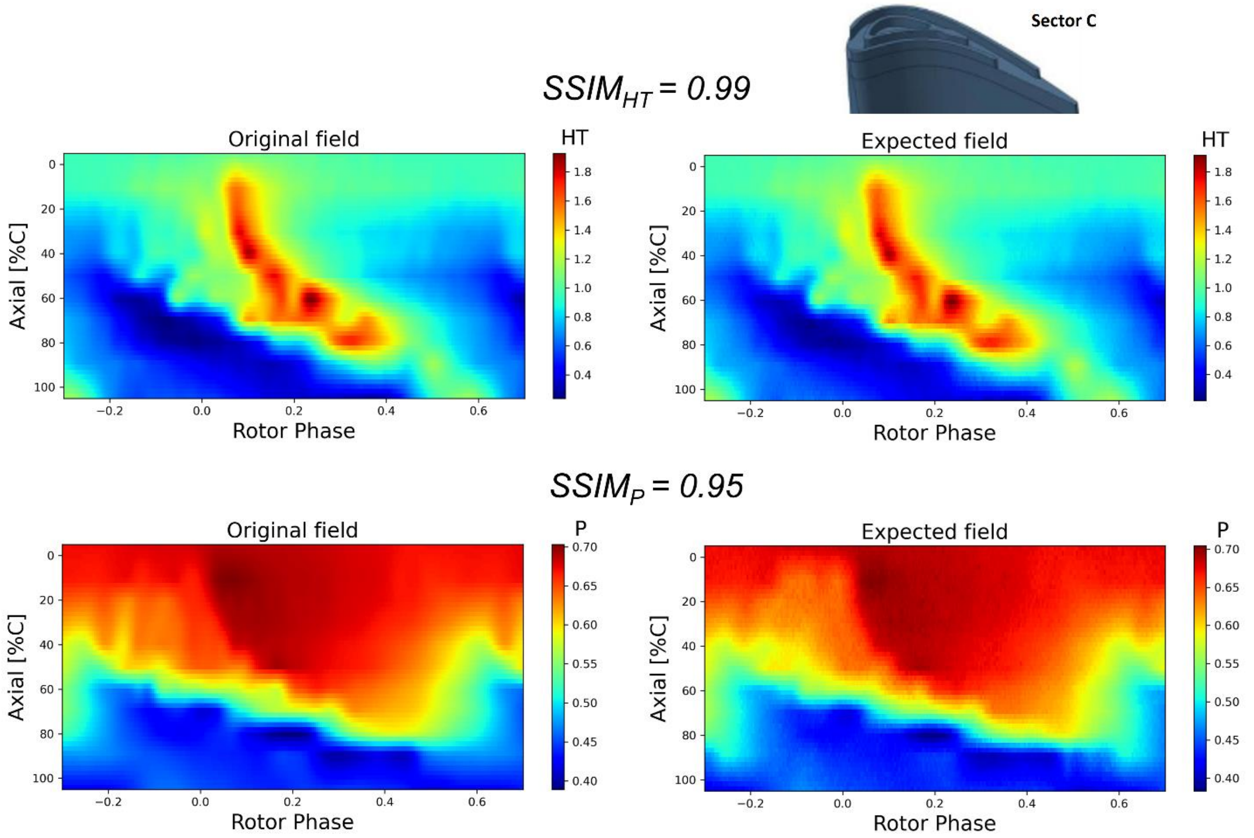


Figure 10 - Example of SSIM evaluation between the original and expected fields of both heat transfer and pressure for sector C.

A value of SSIM=1 means perfect similarity between two images, both in terms of shape and contrast. A quasi-perfect similarity emerges when comparing the original and expected pressure fields, while the similarity is complete for the heat transfer fields (Table 4).

Table 4 - SSIM per sector.

	Sector A	Sector B	Sector C
$SSIM_P$	0.95	0.96	0.95
$SSIM_{HT}$	0.99	0.99	0.99

CONCLUSIONS

In this work an unsupervised dimensionality reduction strategy was proposed. This method is tested against experimental data from a turbine test rig at the von Karman Institute, which has 48 blades separated into 7 sectors with the same baseline rotor geometry but various tip contouring geometries.

The aim of the framework is to unveil causal relationships between tip geometry and unsteady aero-thermodynamic performance, exploiting methods commonly applied to image and sound recognition. The models first up-sample geometries and performance data and then perform unsupervised dimensionality reduction through PCA e AEs to project the original data on latent spaces. Latent space is eventually treated using regression algorithms for further processing in turbomachinery design and optimization.

The resulting maps are in excellent agreement with the original data. A quasi-perfect similarity ($SSIM_{HT}$=0.99 and $SSIM_{P}$=0.96), evaluated using similarity algorithms, is achieved for both pressure and heat transfer fields.

REFERENCES

Akyurt, İbrahim Zeki, Yusuf Kuvvetli, and Muhammet Deveci. (2020). *Enterprise Resource Planning in the Age of Industry 4.0: A General Overview.* Logistics 4.0: 178-185.

Alic, A. S., Almeida, J., Aloisio, G., Andrade, N., Antunes, N., Ardagna, D., ... & Vieira, M. (2019). *BIGSEA: A Big Data analytics platform for public transportation information*. Future generation computer systems, 96, 243-269.

Angelini, G., Corsini, A., & Lavagnoli, S. (2021). *Machine-learnt topology of complex tip geometries in gas turbine rotors*. Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy, 235(3), 383-392.

Baldi, P. (2012). *Autoencoders, unsupervised learning, and deep architectures*. In Proceedings of ICML workshop on unsupervised and transfer learning (pp. 37-49). JMLR Workshop and Conference Proceedings.

Bergstra, J., & Bengio, Y. (2012). *Random search for hyper-parameter optimization*. Journal of machine learning research, 13(2).

Biedermann, T. M., Kameier, F., & Paschereit, C. O. (2019). *Successive aeroacoustic transfer of leading-edge serrations from single airfoil to low-pressure fan application*. Journal of Engineering for Gas Turbines and Power, 141(10).

Boursier Niutta, Wehrle, E. J., Duddeck, F., & Belingardi, G. (2018). *Surrogate modeling in design optimization of structures with discontinuous responses: A new approach for ill-posed problems in crashworthiness design*. Structural and Multidisciplinary Optimization, 57(5), 1857–1869. https://doi.org/10.1007/s00158-018-1958-7

Cernat, B. C., Pátý, M., De Maesschalck, C., & Lavagnoli, S. (2019). *Experimental and numerical investigation of optimized blade tip shapes—Part I: Turbine rainbow rotor testing and numerical methods*. Journal of Turbomachinery, 141(1), 011006.

Colmenares-Quintero, R. F., Quiroga-Parra, D. J., Rojas, N., Stansfield, K. E., & Colmenares-Quintero, J. C. (2021). *Big Data analytics in Smart Grids for renewable energy networks: Systematic review of information and communication technology tools*. Cogent Engineering, 8(1), 1935410.

Corsini, A., Delibra, G., Tieghi, L., & Tucci, F. A. 2021. *Cascade With Sinusoidal Leading Edges: Identification And Quantification of Deflection With Unsupervised Machine Learning*. In Turbo Expo: Power for Land, Sea, and Air (Vol. 84898, p. V001T10A006). American Society of Mechanical Engineers.

Daily, J., & Peterson, J. (2017). *Predictive maintenance: How big data analysis can improve maintenance*. In Supply chain integration challenges in commercial aerospace (pp. 267-278). Springer, Cham.

Karatas, M., Karacan, I., & Tozan, H. (2018). *An integrated multi-criteria decision-making methodology for health technology assessment*. European Journal of Industrial Engineering, 12(4), 504-534.

Lavagnoli, S., Paniagua, G., Tulkens, M., & Steiner, A. (2012). *High-fidelity rotor gap measurements in a short-duration turbine rig*. Mechanical Systems and Signal Processing, 27, 590-603.

Majeed, A., Zhang, Y., Ren, S., Lv, J., Peng, T., Waqar, S., & Yin, E. (2021). *A big data-driven framework for sustainable and smart additive manufacturing*. Robotics and Computer-Integrated Manufacturing, 67, 102026.

Miele, E. S., Bonacina, F., & Corsini, A. (2022). *Deep anomaly detection in horizontal axis wind turbines using graph convolutional autoencoders for multivariate time series*. Energy and AI, 8, 100145.

Moreira, A., & Santos, M. Y. (2007). *Concave hull: A k-nearest neighbours approach for the computation of the region occupied by a set of points*.

Pongetti, J., Kipouros, T., Emmanuelli, M., Ahlfeld, R., & Shahpar, S. (2021). *Using Autoencoders and Output Consolidation to Improve Machine Learning Models for Turbomachinery Applications*. In Turbo Expo: Power for Land, Sea, and Air (Vol. 84935, p. V02DT36A018).

Raj, A., Dwivedi, G., Sharma, A., de Sousa Jabbour, A. B. L., & Rajak, S. (2020). *Barriers to the adoption of industry 4.0 technologies in the manufacturing sector: An inter-country comparative perspective*. International Journal of Production Economics, 224, 107546.

Russom, P. (2011). *Big data analytics*. TDWI best practices report, fourth quarter, 19(4), 1-34.

Singh, H. (2021). *Big data, industry 4.0 and cyber-physical systems integration: A smart industry context*. Materials Today: Proceedings, 46, 157-162.

Tieghi, L., Corsini, A., Delibra, G., & Tucci, F. A. (2021). *A Machine-Learnt Wall Function for Rotating Diffusers*. Journal of Turbomachinery, 143(8).

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). *Image quality assessment: from error visibility to structural similarity*. IEEE transactions on image processing, 13(4), 600-612.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2013). *Data mining with big data*. IEEE transactions on knowledge and data engineering, 26(1), 97-107.

Xie, R., Chen, M., Liu, W., Jian, H., & Shi, Y. (2021). *Digital twin technologies for turbomachinery in a life cycle perspective: a review*. Sustainability, 13(5), 2495.

Yang, Y., & Song, X. (2022). *Research on face intelligent perception technology integrating deep learning under different illumination intensities*. Journal of Computational and Cognitive Engineering, 32-36.

Zalger, J. (2017). *Application of variational autoencoders for aircraft turbomachinery design*. Technical report.

Zhang, D., Shi, W., Van Esch, B. B., Shi, L., & Dubuisson, M. (2015). *Numerical and experimental investigation of tip leakage vortex trajectory and dynamics in an axial flow pump*. Computers & fluids, 112, 61-71.

Zhao, H., Yuen, P. C., & Kwok, J. T. (2006). *A novel incremental principal component analysis and its application for face recognition*. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 36(4), 873-886.