

Thermal Imaging-Based Defects Prediction in High-Pressure Die Casting Using Hybrid Neural Networks and Fuzzy Cognitive Maps^{*}

T. Michno¹ **, R. Holom², S. Schmalzer², P. Meyer-Heye¹, G. Scampone^{1,3}, E. Riegler^{1,3}, M. Hartmann^{1,3}, U. Repanšek⁴, N. Košir⁴, P. Šifrer⁴, and K. Poczeta^{5**}

¹ Austrian Institute of Technology GmbH, Austria

² RISC Software GmbH, Austria

³ LKR Light Metals Technologies, Austria

⁴ LTH Castings d.o.o., Slovenia

⁵ Kielce University of Technology, Poland

Abstract. Producing a defect-free, lightweight, high-performance and complex geometry metal components is a highly challenging task. In this paper, we focused on High Pressure Die Casting (HPDC), proposing a hybrid AI model for non-destructive, in-line, and non-process-interrupting defect prediction, using thermal images. For that, a deep neural network model is used to extract features, which are then classified by a Fuzzy Cognitive Map (FCM). Experimental results show that the method improves prediction performance.

The main contributions of this research include: (i) a novel hybrid model architecture for processing thermal images, (ii) a feature extractor for a FCM-based classifier, (iii) extension of FCM via three clustering techniques to enhance classification accuracy, (iv) a modular design, allowing easy addition of other data sources and classes without retraining, (v) a thorough evaluation through model comparisons and an ablation study, and (vi) to the best of our knowledge, first usage of FCM for this problem.

Keywords: HPDC, defect detection, Fuzzy Cognitive Maps, Thermal imaging, Hybrid AI, Industry 4.0.

1 Introduction

Nowadays, demand for lightweight, high-performance, and complex geometry components is growing rapidly, e.g., due to the enriching automotive or aerospace industry. However, obtaining high-quality elements is a very tough process with many challenges. The metaFacturing project, funded by the Horizon Europe program, concentrates on some of them, preparing a digitalized tool-chain with a focus on reduction of operator cost and effort, as well as waste, and incorporating more sustainable use of raw materials. The project considers two use-cases: welding and casting.

One of the problems in quality assurance in High Pressure Die Casting (HPDC) are time-consuming parts tests and destructive nature of some of them. However, for preliminary evaluation, in order to reduce the number of parts to examine, some inline, non-destructive methods can be used. For example, process parameters or thermal images can be used, which is described more in the next section.

Despite the current advancement in applying machine learning and AI to HPDC area and thermal image acquisition methods, there are still some challenges. Acquiring proper

^{*} The work described in this paper is supported by the metaFacturing project (GA 101091635), which has received funding under the Horizon Europe programme.

^{**} Corresponding author. Email: Tomasz.Michno@ait.ac.at.

^{***} Fuzzy Cognitive Maps volunteer consultant, not supported by the metaFacturing project (GA 101091635) and not a part of the consortium. Work limited to consulting the concepts of FCM, without access to any of the project's IP.

thermal data is problematic due to high influence of external factors like e.g. reflections or additional light source on the reliability. Additionally, proper camera calibration is crucial. Another problem which most often occurs is the imbalance of dataset used for training, when using real production samples.

In this paper, we focus on HPDC, proposing a hybrid AI model for non-destructive defects prediction in produced parts, which can be applied in-line, without process interruption and waste production. For that purpose, thermal imaging is incorporated, using images before and after mold spraying. All images are processed by a deep neural model in order to extract features, which are then classified using Fuzzy Cognitive Map (FCM).

The motivation for this work was a need of explainable and scalable HPDC defect prediction classifier, which can be deployed in a real manufacturing environment. The proposed approach tries to overcome limitations of existing methods by incorporating real-life data both for training and testing, as well as providing modularity and scalability, allowing for the easy addition of new data sources and classes without requiring retraining. Additionally, explainability is improved through the FCM utilization, enhancing the decision-making for the Industry 4.0 applications.

The key contributions of this paper are as follows:

- We propose a novel hybrid model architecture, that combines deep thermal images features extraction with interpretable reasoning using FCM, with probabilistic output for each class and a graph structure, where two images (before and after spraying) are simultaneously processed.
- We extend the FCM with three different types of clustering methods (Fuzzy c-Means, Gaussian Mixture Models, and Spectral Clustering) for the training and inference phases, showing significant improvements in classification performance.
- The modular nature of the proposed architecture allows easy integration of additional data sources (e.g., sensor readings from the HPDC machine) and new classes, without retraining requirement of the existing model components.
- We conduct a comprehensive performance comparison of various models and architectures, along with an ablation study to assess the contribution of each component in the proposed architecture.
- To the best of our knowledge, this work is the first one applying FCM-based classifier to in-line HPDC defect prediction.
- Additionally, this research contributes to the advancement of interpretable AI methods in manufacturing, especially to real applications for Industry 4.0; for both training and testing real production data has been used.

Through this work, which is part of a larger research effort, we aim to address the challenges of HPDC defect prediction using thermal images. This represents the first step toward building a more advanced, combined classifier that incorporates multiple data sources available in-line.

This paper is organized as follows: in the next section, related research is discussed. Then, our methodology and solution are presented. Section 4 contains the experimental results along with their discussion, while Section 5 extends these results through an ablation study. The final section provides the conclusion and our directions for future work.

2 Related Work

2.1 Defects occurring in HPDC

HPDC is a highly complex manufacturing process where the final quality of cast components is influenced by numerous process parameters. A careful control of the melt temper-

ature and process parameters (e.g., injection velocity, applied pressure, shot sleeve temperature, use of vacuum, etc.) is essential to optimize the mechanical and microstructural properties of the cast parts [17]. Additionally, the thermal regulation of the die and the mold surface temperature are crucial factors to ensure the production of high-quality components [16].

The improper selection of melt and mold temperatures, along with the incorrect casting parameters, can result in the formation of severe defects [20]. These defects are typically categorized based on their location as either surface or internal defects [7]. Surface defects - such as cold flows, ejection marks, and incomplete filling - are generally detected through visual inspection. In contrast, internal porosities require more advanced inspection techniques. Shrinkage and gas porosities are among the most common internal defects in HPDC components. While improper injection parameters and inadequate use of vacuum-assisted systems are primary contributors to their formation, improper thermal die layout and unsuitable spraying conditions can also significantly influence their occurrence. Internal defects are commonly detected through X-ray imaging, computed tomography (CT), leakage tests, or destructive methods [21]. However, these techniques are often time-consuming and not suitable for in-line or on-site application during the casting process [21]. In contrast, real-time monitoring of die temperature using thermal cameras offers a fast, in-line method for preliminary quality assessment during casting [21]. Although consistent thermal images do not guarantee defect-free castings, monitoring die surface temperature can aid in controlling defect formation associated with hot spots or poor thermal management during the casting process.

2.2 Computer-aided Defects Prediction in HPDC

Due to the latest advances in computer vision, deep learning, and AI, computer-aided defect prediction for HPDC is becoming increasingly popular, providing more and more precise results.

Various methods are used in HPDC defect prediction. For example, Gupta et al. proposed an ensemble model with ResNet50 and CNN for quality analysis of casted samples [10]. The ensembling technique helps with model overfitting, which is a common problem [10]. The most similar work to ours is [15], where the Hypothesis Pruning Generative Adversarial Network (HP-GAN) is used to predict porosity, combining line data from the process and thermal images of the mold before and after spraying. The presented method seems to have high performance; however, only the AUC metric and confusion matrices were used. The results cannot be compared to our method due to the different mold used and the lack of many details about the method and dataset. Additionally, contrary to our solution, the method in [15] is, due to its architecture, not easily scalable to different defect types and data sources. Lastly, our approach decouples feature extraction from the final classification using FCM, providing explainability and modularity, which is not addressed by [15] work.

Furthermore, for porosity prediction, methods using simulations with a 1D diffusion-controlled model are incorporated [9]. Additionally, other defect types can also be predicted using simulations, such as Cold Shuts and Misruns [14]. For simulations, specific software like ANSYS Fluent, Anycasting, ProCAST, or FLOW-3D is sometimes used.

There is also research merging different data gathered using, e.g., the Internet of Things, NoSQL storage, and data analysis tools (e.g., Spark). Additionally, to make predictions, different machine learning techniques based on sensor data are used, such as random forests or neural networks [19]. Using thermal images and deep learning has also been successfully applied in other areas than HPDC, e.g. for porosity prediction and in-situ monitoring, such

as laser powder bed fusion [23]. For non-destructive defect detection, X-ray can be used, especially in connection with deep learning (DL)-based classifiers [6], such as CNN and ResNet18 [32].

2.3 Fuzzy Cognitive Maps

Fuzzy Cognitive Maps (FCM) are used in many areas where interpretable knowledge representation is needed [34]. They have been applied with great success in the industrial and engineering fields [24], e.g., for autonomous collision avoidance at sea [8], gas prediction [25] or time series data analysis [26]. Recently, they have started being used with very promising results for image classification [12, 30] or defect inspection using cameras [34]. The main idea is based on feature extraction from the image, using, e.g., the VGG-16 network [35] or Vision Transformer [31], followed by the creation of a fuzzy cognitive map based on information granules [35]. Not only binary classification [34] but also multi-label classification is supported [31].

3 Methodology

3.1 Data Acquisition

Data for this research was obtained using the real production line at one of the project partner's factory. Varying casting parameters were used to monitor the process and form defects. For each die casting shot, different data were collected:

1. Structured process control parameters and measurements made by the HPDC machine, such as maximum metal pressure; additionally, information from the leakage tests is gathered at the end of production process,
2. Configured process parameters,
3. Time series data from the actual casting process,
4. Thermal images obtained using an infrared camera pointed to the casting tool surface,
5. Quality checks made by machine operators.

In order to determine part quality, the following combination of methods was used: automated quality checks made by the HPDC machine, visual inspection, and leakage tests (which provide the most valuable feedback on overall quality).

Following the projects data management plan assured reliable data storage and integration of the different data sources. Future plans include e.g. complying with machine-actionable data management plans as suggested in [22].

3.2 Architecture Overview

The architecture of the system was designed to provide a highly scalable model with the possibility of easily adding new data sources as well as new classes. One of the solutions that allows such flexibility is FCM [18], with some extensions, such as for image classification [12, 30, 34]. The model consists of two images of the mold: before and after spraying, which are passed to the feature extraction module based on a Deep Neural Network (DNN) architecture. After feature extraction, the features are then used to build a FCM, which, during the inference process, provides the final classification. As a result, for the use case considered in this paper, only two classes are predicted, with values between 0 and 1. However, the proposed architecture is highly flexible and more classes can be added, even without retraining. As it can be seen, unlike other methods described in previous section,

our approach decouples feature extraction from the final classification using FCM, providing explainability and modularity. An overview of the solution architecture is shown in Fig. 1.

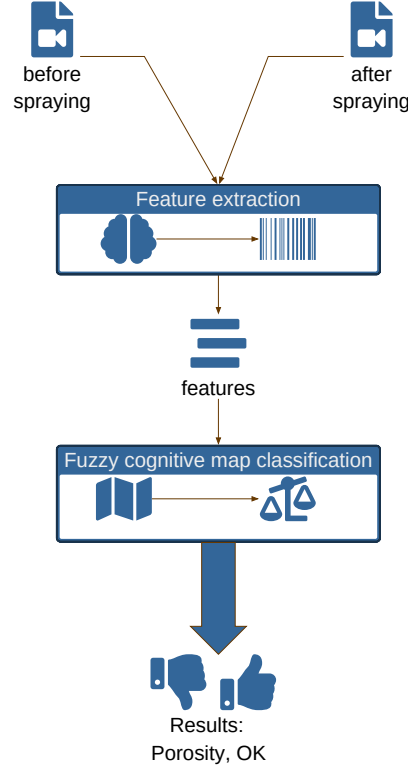


Fig. 1: General overview of the solution architecture.

Feature Extraction - Deep Neural Networks One of the most common approaches for image classification nowadays is to use a DNN architecture. This is very efficient, especially when connected with autoencoders [3]. In this paper, a similar approach has been chosen, based on a DNN and an encoder extracted from a trained autoencoder model. Additionally, because there are two images for each sample, one before and one after spraying, the DNN and encoder branches are duplicated.

For the DNN branch, state-of-the-art and high-performance architectures have been chosen, such as ResNet, VGG, and DenseNet [5]. Since deep models require large training datasets, pretrained models can be used with additional retraining on domain-specific data [27], which in our case are thermal images.

The other part of the feature extractor uses an encoder, extracted from an autoencoder network trained on thermal images. Since the images before and after spraying have slightly different characteristics (e.g., temperature ranges), two trained models have been created. For the autoencoder architecture, a U-Net [36]-inspired model has been chosen. After training with images, to obtain a feature vector from the neuron outputs, the last layers are removed, forming, for each i -th sample (before and after images), a feature vector $\mathbf{f}_i$ consisting of x_j float values, defined as follows:

$$\mathbf{f}_i = [x_1, x_2, \dots, x_d], \quad \text{where } x_j \in \mathbb{R} \quad (1)$$

Fuzzy Cognitive Maps A FCM is a type of single-layer recurrent neural network. The structure of the FCM model is based on a directed graph whose nodes denote concepts important for the analyzed system. Edges describe the causal relationships between concepts. Weights of the relationships between concepts can be determined by experts or based on available data. Concept values are initialized based on the initial vector and then updated iteratively until the system converges or a maximum number of iterations is reached. The concepts values can be updated with the use of the following inference rule [18]:

$$X_i(t+1) = f\left(\sum_{j=1, j \neq i}^n w_{j,i} X_j(t)\right), \quad (2)$$

where:

- $X_i(t)$ is the value of the i -th concept,
- t is the number of iteration,
- $w_{j,i}$ is the weight of the relationships between the j -th concept and the i -th concept,
- $i, j = 1, \dots, n$, n is the number of concepts,
- $f(x)$ is a sigmoid function that normalizes the concept values to the interval $[0,1]$ [4].

The main idea of image classification using FCM in our research has been based on the work by Tziolas et al. [34]. The features extracted from the images are clustered using the K-means algorithm. The input concepts of the FCM model are initialized based on the centroid of each cluster. The output concepts are equal to the number of analyzed classes. The weights of the relationships between concepts are determined based on the fuzzy similarities between the extracted feature vectors and clusters' centroids.

In our research, in order to improve the results as well as refine the method to our use-case, we proposed to additionally use three different types of cluster centroids: Fuzzy c-Means [2], Gaussian Mixture Model [29] and Spectral clustering [1].

In the Fuzzy c-Means algorithm, for each i -th feature vector $\mathbf{f}_i$ a set of coefficients $w_{i,j}$ is computed, which describes the degree of belonging to the j -th cluster. Afterward, a j -th centroid $\mathbf{c}_{fcm(j)}$ is computed. In the case of Gaussian Mixture Model (GMM)-based k -th centroid $\mathbf{c}_{gmm(k)}$ computation, each feature vector $\mathbf{f}_i$ is associated with the posterior probability $\gamma_{i,k}$ of belonging to the k -th cluster. For Spectral Clustering, the centroid $\mathbf{c}_{spec(l)}$ computation uses similarity matrix and dimensionality reduction with eigenvectors in order to perform the final clustering in the reduced spectral space. To summarize, all of the centroids can be computed as follows:

$$\mathbf{c}_{fcm(j)} = \frac{\sum_{i=1}^N w_{i,j}^m \mathbf{f}_i}{\sum_{i=1}^N w_{i,j}^m} \quad (3)$$

$$\mathbf{c}_{gmm(k)} = \frac{\sum_{i=1}^N \gamma_{i,k} \mathbf{f}_i}{\sum_{i=1}^N \gamma_{i,k}} \quad (4)$$

$$\mathbf{c}_{spec(l)} = \frac{1}{|\mathcal{C}_l|} \sum_{i \in \mathcal{C}_l} \mathbf{f}_i \quad (5)$$

where:

- $\mathbf{f}_i \in \mathbb{R}^d$ – the i -th feature vector,
- $w_{i,j} \in [0, 1]$ – the membership degree of $\mathbf{f}_i$ to cluster j ,
- $m > 1$ – the fuzziness parameter (typically $m = 2$),
- N – the total number of feature vectors,

- $\gamma_{i,k} \in [0, 1]$ - the posterior probability that $\mathbf{f}_i$ belongs to the k -th cluster,
- $\mathcal{C}_l$ – the set of indices of feature vectors assigned to the l -th cluster,
- $|\mathcal{C}_l|$ – the number of feature vectors in cluster l .

For each clustering method, the shapes of the resulting clusters differ, which contributes to better overall differentiation between image classes. Additionally, the optimal number of centroids for each type is determined by iteratively evaluating model performance using metrics such as the Matthews Correlation Coefficient (MCC) and Balanced Accuracy. For the use-case considered in this paper, the highest performance metrics have been achieved with $j=4$, $k=5$ and $l=5$, corresponding to the number of centroids for Fuzzy c-Means, Gaussian Mixture Model, and Spectral Clustering, respectively.

As a result of the FCM classification usage, interpretable outputs are generated, providing a class-wise certainty level, which helps explainable decision-making.

3.3 Potential for Improvements and Extensions

The proposed solution can will be improved as a part of further research. The areas which are considered are: enhancing the feature extractor architecture (e.g. adding GAN discriminator block), clustering algorithm (e.g. incorporating other machine learning methods, like Genetic algorithms) and features reduction (e.g using PCA).

Additionally, the classifier offers high flexibility and can be easily extended by incorporating additional data sources, such as inline process data (e.g., velocity and pressure values, temperatures), to create a fused model and further improve classification performance. Different types of feature extractors can be employed, whose outputs are then integrated into the constructed Fuzzy Cognitive Map. Furthermore, the architecture supports not only the inclusion of new data sources but also the addition of new defect classes, such as Cold shuts or Misruns. Due to space limitations, a detailed description of the extended approach, along with initial results, which are very promising and show improved metric values, will be presented in a separate publication. Another way of extension is to ensemble models for higher accuracy of classification for the same data, which will be investigated as future research.

4 Experiments and Results

In order to evaluate the proposed solution, an experimental application has been written in Python. As a training dataset, 3685 samples (7370 images) of class "OK" and 306 samples (612 images) of class "porosity" have been used. Due to the fact that the dataset was imbalanced, additional augmented training data have been added for the "porosity" class, resulting in additional samples. Moreover, for comparison with the most similar other approach ([15]), a balanced training and testing datasets have been created, with all "porosity" samples and reduced "OK" samples number, to match them. For training the autoencoder, only real images have been used; for training the DNN branch and FCM, both real and augmented images were utilized. As a test dataset, 154 samples of class "OK" (308 images) and 102 samples of "porosity" (204 images) have been used. For performance evaluation, the following metrics were used: Area Under the Curve (AUC), Accuracy, Matthews Correlation Coefficient (MCC), Balanced Accuracy, and F1-scores for each class.

4.1 Results Using Neural Network-Based Classifier

As a first step, different deep neural network architectures have been examined, including DenseNet [13], ResNet50 [11], EfficientNetB0 [33], YOLOv2 [28], and an Encoder (extracted from a trained Autoencoder network with an architecture inspired by U-Net [36]), as well as ResNet50, EfficientNetB0, and YOLOv2 concatenated with the Encoder.

The classification results using proposed hybrid model are shown in Table 1. It can be seen that the best performance was obtained by ResNet50, ConvNeXt, and DenseNet concatenated with the Encoder branch. Example confusion matrices for ConvNeXt and DenseNet are shown in Fig. 3b and Fig. 4b. However, the difference between F1-scores for both classes is high, reaching up to 0.81 for the "OK" class and 0.51 for the "Porosity" class.

The experiments demonstrated that adding the Encoder branch improves overall performance across almost all architectures, particularly for the MCC, which increased from 0.00 to approximately 0.37–0.48. Architectures without the Encoder, such as ResNet50, EfficientNetB0, and YOLOv2, achieved very low metric values, for example AUC around 0.50 and MCC equal to 0.00, indicating that their performance does not surpass random prediction. A confusion matrix for the test data using YOLOv2 is shown in Fig. 2, illustrating this issue. In contrast, the confusion matrices for DenseNet and ConvNeXt (Fig. 3a and Fig. 4a) show that none of the "OK" class samples were misclassified as "Porosity", but still the performance is lower than for the version with added Encoder.

Table 1: Comparison of the performance of different neural network architectures on the test dataset.

Model(s)	AUC	Accuracy	MCC	Balanced accuracy	F1-score OK	F1-score Porosity
Encoder	0.67	0.47	0.47	0.67	0.81	0.50
DenseNet	0.65	0.71	0.46	0.65	0.80	0.47
ResNet50, EfficientNetB0	0.50	0.42	0.00	0.50	0.00	0.59
YOLOv2	0.50	0.58	0.00	0.50	0.74	0.00
DenseNet + Encoder	0.65	0.71	0.45	0.65	0.80	0.46
ResNet50 + Encoder	0.65	0.71	0.43	0.65	0.80	0.48
EfficientNetB0 + Encoder	0.63	0.69	0.39	0.63	0.79	0.43
ConvNeXt + Encoder	0.67	0.72	0.48	0.67	0.81	0.51
YOLOv2 + Encoder	0.61	0.67	0.37	0.61	0.78	0.35

Actual	OK	0	130
	Porosity	0	93
		OK	Porosity
		Predicted	

Fig. 2: Confusion Matrix for the YOLOv2 model.

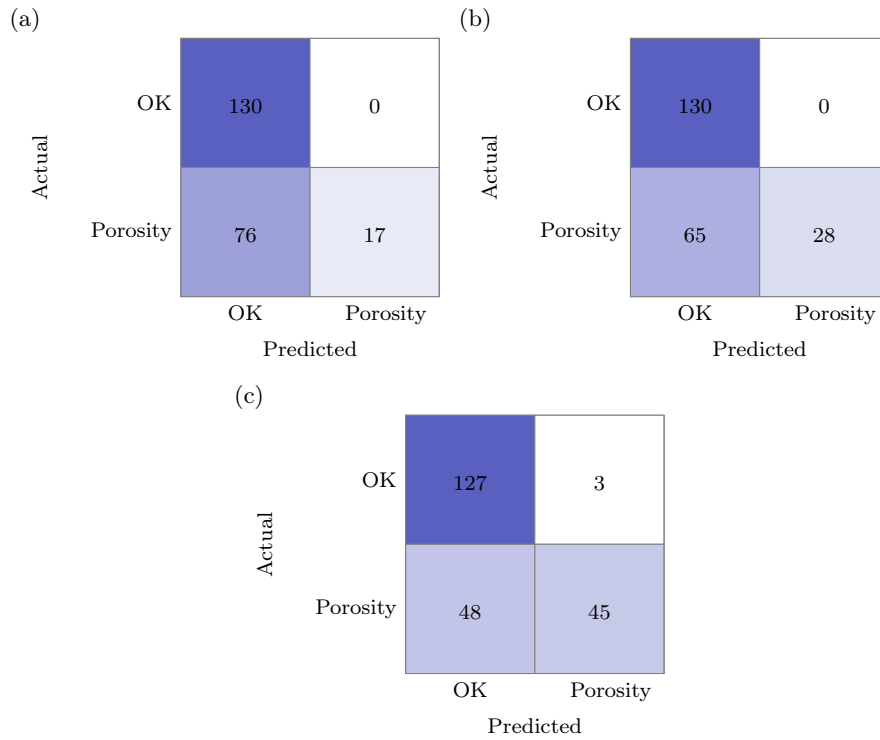


Fig. 3: Confusion matrices for DenseNet variants: (a) DenseNet, (b) DenseNet + Encoder, (c) DenseNet + Encoder + FCM.

4.2 Fuzzy Cognitive Map-Based Classification

The networks trained and evaluated in the previous subsection were used as feature extractors by removing the final layers of the concatenated models. The extracted features were then used to train a FCM. Evaluation was performed using the same images as in the neural network-based classification described in the previous subsection. The experimental results are presented in Table 2.

Some general observations can be made based on the results:

- For models using neural networks as feature extractors and FCM as a classifier, a significant improvement in performance can be observed, especially for those models that include the Encoder branch. In contrast, models without the Encoder achieved the lowest metric values.
- The highest performance was obtained by DenseNet and ConvNeXt models concatenated with the Encoder and subsequently passed to the FCM.

Comparing to the results from the previous subsection, it can be seen that the F1-score for the "Porosity" class obtained much higher values (about 29% higher), additionally still maintaining high values for the "OK" class. The highest F1-score for the "Porosity" class obtained by neural networks was 0.59, however, reducing the F1-score for the "OK" class to 0.00. The highest values for both classes were 0.81 ("OK") and 0.51 ("Porosity"). Contrary to that, the highest obtained values for the FCM classifier were 0.83 ("OK") and 0.64 ("Porosity"), as well as 0.82 ("OK") and 0.66 ("Porosity"). A similar situation can be observed, for example, for the MCC values: for neural networks, the highest one was 0.48, while for the FCM, it was 0.55.

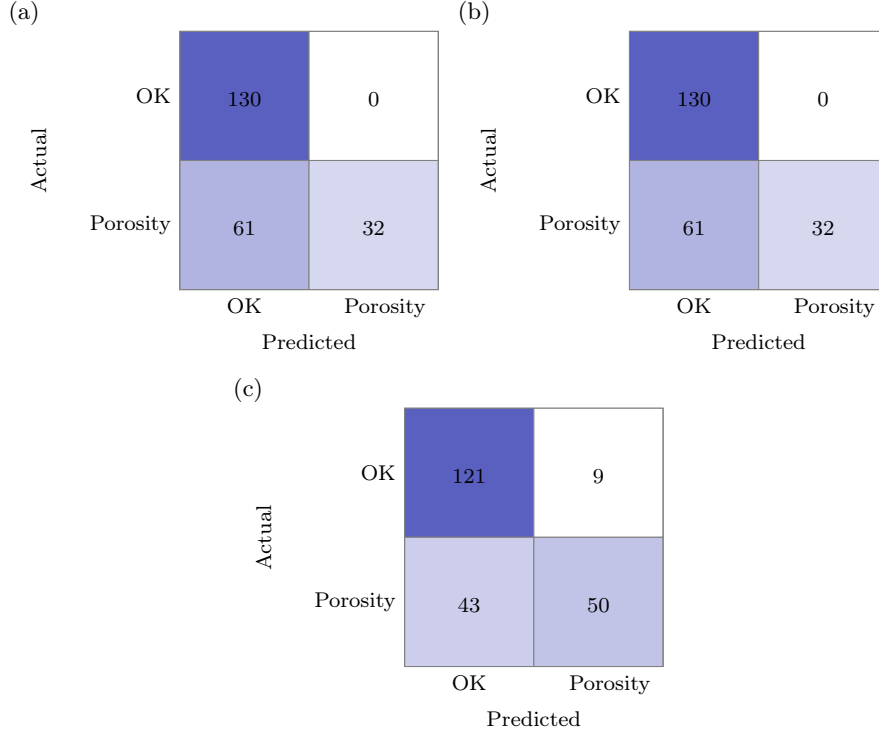


Fig. 4: Confusion matrices for ConvNeXt variants: (a) ConvNeXt, (b) ConvNeXt + Encoder, (c) ConvNeXt + Encoder + FCM.

Comparison of the confusion matrices also shows performance improvement, especially for the "Porosity" class. However, still not all samples can be predicted, but this may be connected to the different origins of the porosity, its severity or location. This situation will be examined in further research on this topic.

As an additional performance test, a comparison with the results of a sensor process data values trained with LightGBM classifier was done. Fig. 7 shows the best FCM models' metric values compared to the non-image-based classifier. It can be observed that the best FCM models are obtaining similar metric values to the LightGBM classifier, with the exception of the MCC, which is only slightly lower for FCM models. However, when comparing the confusion matrices, shown in Fig. 5 and Fig. 3c, the values are almost identical. This means that using two different data sources separately results in similar prediction performance. This situation can be used to combine predictions into one, stronger classifier, which will be investigated further with more details. However, first results are very promising, as can be seen in Fig. 6.

4.3 Comparison with HP-GAN algorithm

In the previous subsection, we presented a comparison between the proposed solution and several state of the art deep neural network architectures, with and without added Encoder branch. As can be seen, our method achieved higher overall performance.

Additional tests have been performed in order to compare our solution with other, more complex researches than the network architectures shown in the previous subsections. As the main algorithm, a HP-GAN described by [15] has been chosen, since it is the most similar research. The implementation has been limited to use only images, as our solution

Table 2: Comparison of the performance of different neural network architectures on the test dataset.

Model(s)	AUC	Accuracy	MCC	Balanced acc.	F1 OK	F1 Porosity
Encoder FCM	0.59	0.65	0.33	0.59	0.77	0.29
ResNet50 FCM	0.50	0.58	0.00	0.50	0.74	0.00
EfficientNetB0 FCM	0.50	0.58	0.00	0.50	0.74	0.00
YOLO FCM	0.50	0.58	0.00	0.50	0.74	0.00
DenseNet + Encoder FCM	0.73	0.77	0.55	0.73	0.83	0.64
ResNet50 + Encoder FCM	0.59	0.56	0.19	0.59	0.52	0.59
EfficientNetB0 + Encoder FCM	0.59	0.56	0.19	0.59	0.50	0.60
ConvNeXt + Encoder FCM	0.73	0.77	0.52	0.73	0.82	0.66
YOLO + Encoder FCM	0.54	0.52	0.09	0.54	0.50	0.54

Actual	OK	129	1
	Porosity	47	46
		OK	Porosity
		Predicted	

Fig. 5: Confusion Matrix for the LightGBM classifier using sensors data.

Actual	OK	100	30
	Porosity	25	68
		OK	Porosity
		Predicted	

Fig. 6: Confusion Matrix for the combined model using both image and sensors data.

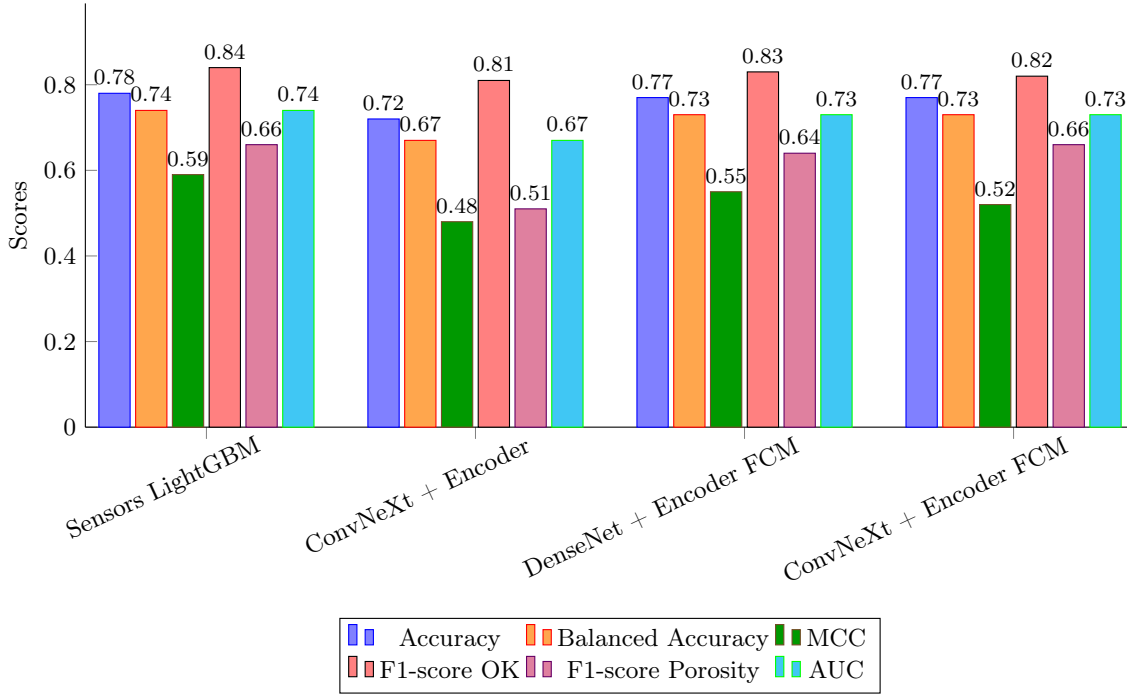


Fig. 7: Comparison of the best model performance metrics.

also is based on them. The tests consisted of training and testing with imbalanced datasets and balanced dataset for both training and testing.

The test results using imbalanced datasets are shown in Table 3. It can be seen that only the AUC and F1-score for "OK" class were higher for HP-GAN, whereas all other metrics were higher for our solution. The performance difference is shown by F1-score and recall metrics for the "Porosity" class, indicating much higher ability of our solution to distinguish the defect from the non-defect samples. Additionally, computation time has been measured: the training took 5060.87s for HP-GAN and 681.59s for our solution, the testing 9.86s for HP-GAN and 8.78s for our solution accordingly.

The test results using imbalanced datasets are shown in Table 4. It can be seen that for the AUC and MCC metrics, the HP-GAN achieved higher results. However, for all others, our solution gave better performance, especially for the Balanced Accuracy. It must be noted that the training dataset was limited to only about 1000 of samples, thus the results may differ, when more data will be captured during the further project development. The computation times were as follows: for the training: 811.04s for HP-GAN and 89.61s for our solution, for the testing: 3.01s for HP-GAN and 8.65s for our solution.

Table 3: Comparison of model performance: HP-GAN [15] vs. our approach for imbalanced dataset.

Model	AUC	MCC	Balanced Acc.	F1 OK	F1 Porosity	Recall OK	Recall Porosity
HP-GAN [15]	0.746	0.313	0.569	0.911	0.244	0.996	0.142
Proposed solution	0.698	0.357	0.698	0.767	0.556	0.690	0.707

Table 4: Comparison of model performance: HP-GAN [15] vs. our approach for balanced dataset.

Model	AUC	MCC	Balanced Acc.	F1 OK	F1 Porosity	Recall OK	Recall Porosity
HP-GAN [15]	0.869	0.820	0.663	0.847	0.794	0.937	0.704
Proposed solution	0.834	0.692	0.834	0.860	0.810	0.954	0.714

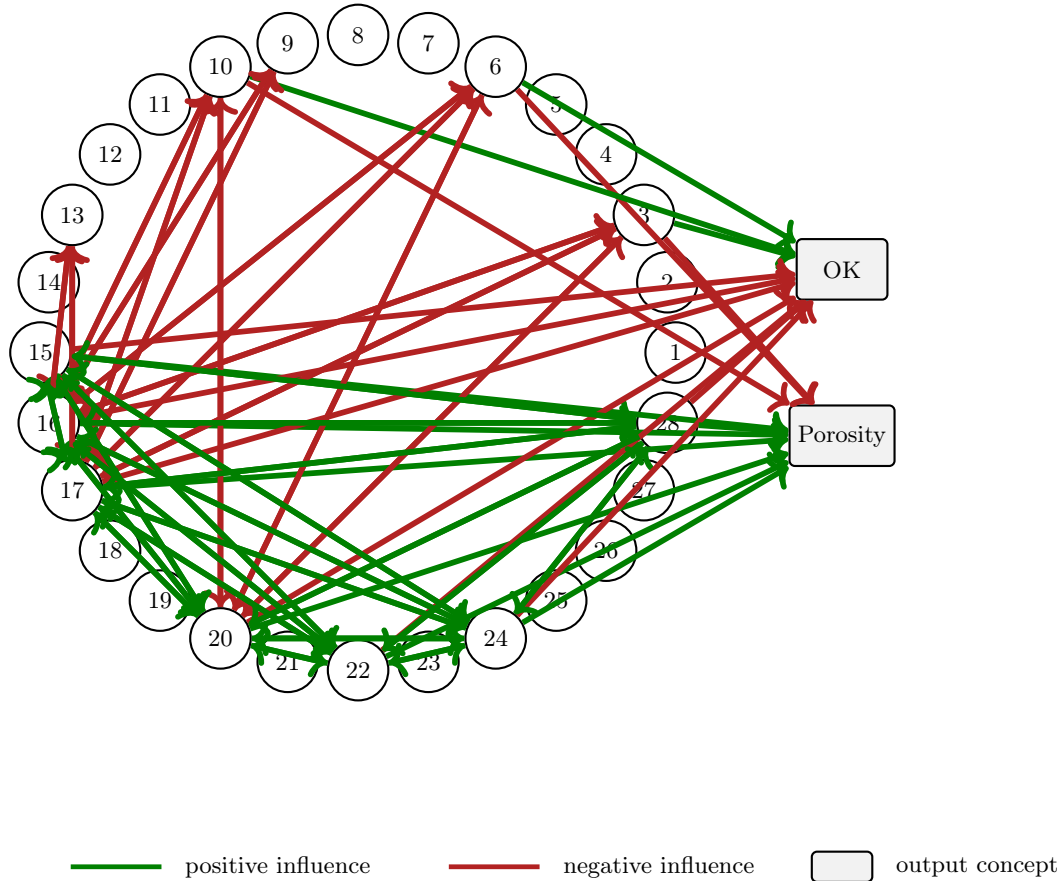


Fig. 8: Example trained FCM visualization, showing global top 10% edges.

4.4 Fuzzy Cognitive Map

An example FCM visualization is shown in the Fig. 8. Due to the high number of edges, only top 10% is shown. However, even with such a limit, it can be seen that some of the input concepts are positively influencing the Porosity class, whereas much less are related to the OK class.

The output classification of the FCM can provide important information for the e.g. machine operator, showing the probability of Porosity occurrence - when it's close to 1.0, then there is high chance that the part will contain it. However, when the value is lower, e.g. closer to 0.5, than additional tests may be needed to be done. An example plot showing the predictions is shown in Fig. 9. Such a plot can be used to see the process stability, when the probability of the class "Porosity" is becoming higher than the probability for "Good", then the process should be test for different problems which may cause the higher number of parts with the porosity.

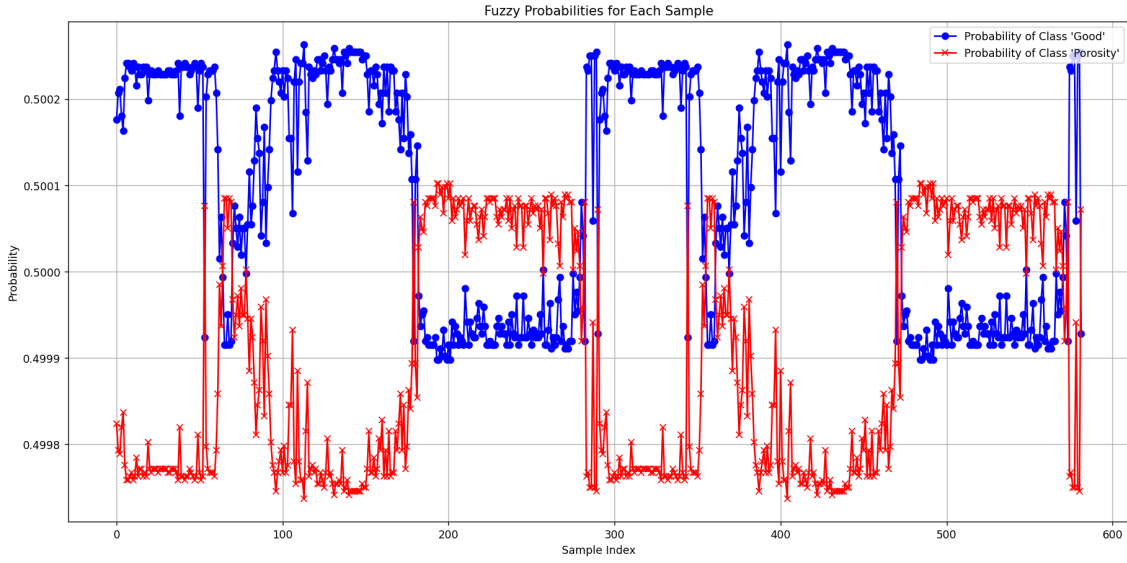


Fig. 9: A visualization of the probabilities for the sample test set.

4.5 Ablation Study

In order to better understand the influence of all components in the proposed architecture, an ablation study was additionally performed. It was divided into two parts: evaluating the feature extractor and evaluating the FCM classifier.

The first part was primarily covered by the experiments performed in subsection 4.1, where different architectures were analyzed. As a result, shown in Table 1 and Table 2, it can be concluded that for both, the feature extractor and FCM, adding the Encoder branch is very beneficial, improving overall performance. Additionally, using FCM stabilizes the model's performance, especially for the "Porosity" class.

Table 5: Comparison of FCM model performance on the test dataset with removed centroids.

Removed type of centroids	AUC	Accuracy	MCC	Balanced accuracy	F1 OK	F1 Porosity
none	0.73	0.77	0.55	0.73	0.83	0.64
Fuzzy c-Means	0.69	0.74	0.50	0.69	0.81	0.56
GMM only	0.60	0.58	0.21	0.60	0.55	0.60
Spectral only	0.67	0.73	0.47	0.67	0.81	0.53
Fuzzy c-Means and GMM	0.56	0.56	0.13	0.56	0.58	0.53
Fuzzy c-Means and Spectral	0.56	0.63	0.27	0.56	0.76	0.21
GMM and Spectral	0.69	0.73	0.46	0.69	0.80	0.56

The second part of the ablation study focused on the FCM and the influence of centroids on the overall performance, which is shown in Fig. 10. Firstly, different numbers of centroids were removed, showing that the model is quite stable up to 5 centroids being lost. After removing 9 centroids, the metric values dropped drastically, and after removing all of them, the system was unable to provide any classification. Further tests were conducted to examine the influence of each centroid type on the results: Fuzzy c-Means, Gaussian Mixture Models (GMM), and Spectral Clustering. The results of the study are shown in Table 5. It can be seen that after removing only Fuzzy c-Means, only Spectral, and

Table 6: Comparison of FCM model performance using different activation functions.

Activation function	AUC	Accuracy	MCC	Balanced accuracy	F1 OK	F1 Porosity
sigmoid ($l=0.65$, $l=0.1$)	0.73	0.77	0.55	0.73	0.83	0.64
sigmoid ($l=0.25$, $l=5$, $l=10$)	0.72	0.76	0.53	0.72	0.83	0.62
softmax	0.72	0.76	0.53	0.72	0.83	0.62
tanh	0.50	0.58	0.00	0.50	0.74	0.00
ReLU	0.50	0.58	0.00	0.50	0.74	0.00

both GMM and Spectral centroids, the AUC, Accuracy, Balanced Accuracy, and F1-scores remain surprisingly high. The lowest MCC value was obtained by removing only GMM, which indicates that this cluster type is the most important for the classification. However, when both GMM and Spectral centroids were removed, the performance did not drop significantly, suggesting that with only Fuzzy c-Means clusters, the classifier was still able to find good matches to the centers. As in the previous experiment, when more clusters were removed, the overall results worsened. Nevertheless, the test showed that all of the centroid types are necessary, and only when they are combined together, the highest performance is achieved. The last tests for the FCM classifier were performed in order to evaluate the influence of choosing different activation functions on the final performance. The results are shown in Table 6 — it can be seen that the highest metric values were obtained using the sigmoid and softmax activation functions. Additionally, it can be noted that the sigmoid function was not very sensitive to the l parameter value.

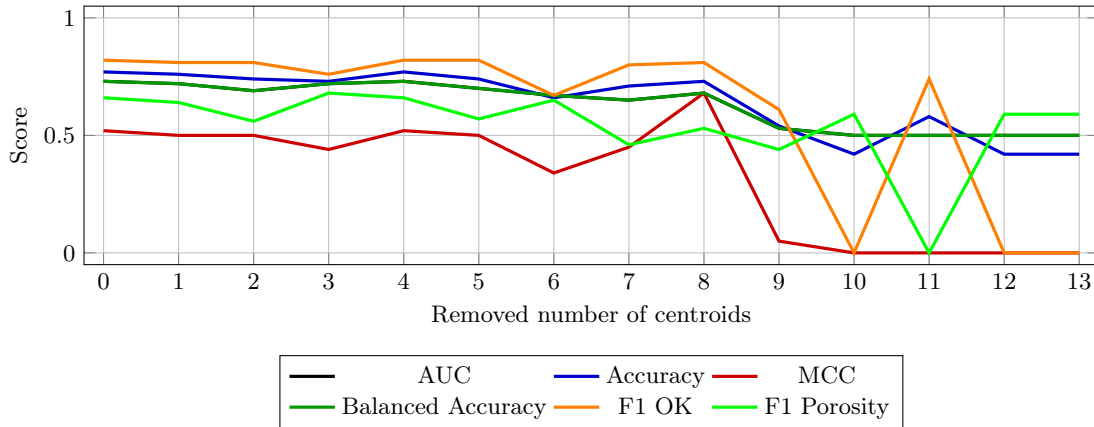


Fig. 10: Comparison of FCM model performance on the test dataset with randomly removed centroids.

4.6 Limitations

The experiments showed the performance of the proposed solution. The results are very promising, however, still there are some limitations found, which can be summarized as follows:

- not all porosity occurrence can be found - which is related to the cause of the defect, as well as its size,
- choosing the right l parameter for sigmoid activation function is important for achieving the highest metrics values,
- the number of chosen centroids also has influence on the performance.

5 Conclusion

In this paper, a novel application of FCM classification to thermal images of the HPDC process has been presented, contributing to trustworthy AI in Industry 4.0 and supporting the incremental system evolution. The classifier, although based on the work by [34], has been improved by introducing a novel feature extractor and incorporating different clustering methods during the training and inference phases. The classification is focused on porosity prediction in casted parts. Compared to other methods, this approach allows the easy addition of new classes as well as additional data sources, without the need to retrain already trained parts of the classifier. The results prove that the proposed solution improves the performance of prediction; however, probably due to the different nature of porosity, its severity or location, not all occurrences can be found. This issue will be further investigated. When comparing to other solutions, the HP-GAN achieved better results for the AUC (balanced and imbalanced datasets) and MCC metrics (only with balanced dataset). However, for all others, especially the Balanced Accuracy, our solution obtained higher values. Additionally, the training time is much lower than for HP-GAN, as well as both training and testing times are not increasing as fast as for HP-GAN, when the dataset size increases. In this paper, we also performed an ablation study, which demonstrates the influence of each part of the classifier on the final performance.

The solution described in this paper is part of an ongoing research on the quality inspection of HPDC parts. As part of future work, further improvements to the classifier's performance will be investigated. One of the methods involves using additional data sources, such as structured process control parameters and measurements data, to supplement the classifier. The initial results are very promising, as shown in Fig. 6. Additionally, different training methods will be explored to prepare weights in a more efficient manner. Ultimately, the classifier will be integrated into a real-life smart manufacturing environment with a dashboard for factory employees as a part of the metaFacturing project.

Acknowledgment

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

The work described in this paper is supported by the metaFacturing project (GA 101091635), which has received funding under the Horizon Europe programme.

References

1. Alpert, C.J., Kahng, A.B., Yao, S.Z.: Spectral partitioning with multiple eigenvectors. *Discrete Applied Mathematics* **90**, 3–26 (1 1999). [https://doi.org/10.1016/S0166-218X\(98\)00083-3](https://doi.org/10.1016/S0166-218X(98)00083-3)
2. Bezdek, J.C.: Pattern recognition with fuzzy objective function algorithms. *Pattern Recognition with Fuzzy Objective Function Algorithms* (1981). <https://doi.org/10.1007/978-1-4757-0450-1>
3. Bhagat, S., Baheti, V., Sinha, M., Pandey, M., Goyal, M.: Deep learning based image classification using auto-encoders. *Proceedings of the 5th International Conference on Data Intelligence and Cognitive Informatics, ICDICI 2024* pp. 697–703 (2024). <https://doi.org/10.1109/ICDICI62993.2024.10810902>
4. Bueno, S., Salmeron, J.L.: Benchmarking main activation functions in fuzzy cognitive maps. *Expert Systems with Applications* **36**(3, Part 1), 5221–5229 (2009). <https://doi.org/https://doi.org/10.1016/j.eswa.2008.06.072>
5. Deepthi, G., Bamini, A.M.A., Praveen, Y.J.: A comparative analysis of pancreatic tumor detection using vgg16, resnet, and densenet. *Proceedings of the 3rd International Conference on Applied Artificial Intelligence and Computing, ICAAIC 2024* pp. 1303–1308 (2024). <https://doi.org/10.1109/ICAAIC60222.2024.10575802>

6. İsmail Enes Parlak, Emel, E.: Deep learning-based detection of aluminum casting defects and their types. *Engineering Applications of Artificial Intelligence* **118**, 105636 (2 2023). <https://doi.org/10.1016/J.ENGAPPAI.2022.105636>
7. Gariboldi, E., Bonollo, F., Rosso, M.: Proposal of a classification of defects of high-pressure diecast products. *La Metallurgia Italiana* (6 2007)
8. Gil, M., Poczęta, K., Wróbel, K., Yang, Z., Chen, P.: Toward using fuzzy grey cognitive maps in manned and autonomous collision avoidance at sea. *IEEE Journal of Oceanic Engineering* pp. 1–21 (2025). <https://doi.org/10.1109/JOE.2024.3516095>
9. Godwal, U., Bhagavath, S., Ghaffari, B., Li, M., Lee, P.D., Karagadde, S.: Numerical modelling of porosity with combined gas and shrinkage effects in hpdc. *IOP Conference Series: Materials Science and Engineering* **1274**, 012028–012028 (1 2023). <https://doi.org/10.1088/1757-899X/1274/1/012028>
10. Gupta, R., Anand, V., Gupta, S., Koundal, D.: Deep learning model for defect analysis in industry using casting images. *Expert Systems with Applications* **232**, 120758 (12 2023). <https://doi.org/10.1016/J.ESWA.2023.120758>
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2016-December**, 770–778 (12 2015). <https://doi.org/10.1109/CVPR.2016.90>
12. Hilal, A.M., Alsolai, H., Al-Wesabi, F.N., Nour, M.K., Motwakel, A., Kumar, A., Yaseen, I., Zamani, A.S.: Fuzzy cognitive maps with bird swarm intelligence optimization-based remote sensing image classification. *Computational Intelligence and Neuroscience* **2022**, 4063354 (1 2022). <https://doi.org/10.1155/2022/4063354>
13. Huang, G., Liu, Z., Maaten, L.V.D., Weinberger, K.Q.: Densely connected convolutional networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* **2017-January**, 2261–2269 (8 2016). <https://doi.org/10.1109/CVPR.2017.243>
14. Jakumeit, J., Behnken, H., Laqua, R., Mbewou, S., Fehlbier, M., Gänz, J., Becker, L.: Coupled modeling of misrun, cold shut, air entrainment, and porosity for high-pressure die casting applications. *Minerals, Metals and Materials Series* **6**, 865–870 (2021). https://doi.org/10.1007/978-3-030-65396-5_114
15. Kim, J., Park, C., Park, W., Park, Y., Cho, C., Kim, D.: A study on high pressure die-casting defect prediction deep learning algorithm for porosity defect detection based on process parameters and thermal image. *CDE* **28**, 222–231 (9 2023). <https://doi.org/10.7315/CDE.2023.222>
16. Kong, L.X., She, F.H., Nahavandi, S., Kouzani, A.Z.: Die temperature monitoring of high pressure die casting. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* **3**, 1756–1761 (2000). <https://doi.org/10.1109/ICSMC.2000.886363>
17. Koru, M., Serçe, O.: The effects of thermal and dynamical parameters and vacuum application on porosity in high-pressure die casting of a383 al-alloy. *International Journal of Metalcasting* **12**, 797–813 (10 2018). <https://doi.org/10.1007/s40962-018-0214-7>
18. Kosko, B.: Fuzzy cognitive maps. *International Journal of Man-Machine Studies* **24**, 65–75 (1 1986). [https://doi.org/10.1016/S0020-7373\(86\)80040-2](https://doi.org/10.1016/S0020-7373(86)80040-2)
19. Lee, J.H., Noh, S.D., Kim, H.J., Kang, Y.S.: Implementation of cyber-physical production systems for quality prediction and operation control in metal casting. *Sensors* 2018, Vol. 18, Page 1428 **18**, 1428 (5 2018). <https://doi.org/10.3390/S18051428>
20. Lee, S., Han, D., Kang, S., Kim, N.: Method of predicting shrinkage defects and deriving process conditions in hpdc (high-pressure die-casting) for electric vehicle motor housings. *International Journal of Metalcasting* **18**, 1262–1272 (4 2024). <https://doi.org/10.1007/s40962-023-01100-y>
21. Maierhofer, C., Myrach, P., Röllig, M., Jonietz, F., Illerhaus, B., Meinel, D., Richter, U., Miksche, R.: Characterization of pores in high pressure die cast aluminum using active thermography and computed tomography. *AIP Conference Proceedings* **1706** (2 2016). <https://doi.org/10.1063/1.4940580/585865>
22. Miksa, T., Simms, S., Mietchen, D., Jones, S.: Ten principles for machine-actionable data management plans. *PLOS Computational Biology* **15**, e1006750 (3 2019). <https://doi.org/10.1371/JOURNAL.PCBI.1006750>
23. Oster, S., Breese, P.P., Ulbricht, A., Mohr, G., Altenburg, S.J.: A deep learning framework for defect prediction based on thermographic in-situ monitoring in laser powder bed fusion. *Journal of Intelligent Manufacturing* **35**, 1687–1706 (4 2024). <https://doi.org/10.1007/S10845-023-02117-0/FIGURES/16>
24. Papageorgiou, E.I. (ed.): *Fuzzy Cognitive Maps for Applied Sciences and Engineering*, vol. 54. Springer Berlin Heidelberg (2014). <https://doi.org/10.1007/978-3-642-39739-4>
25. Poczęta, K., Papageorgiou, E.I.: Implementing fuzzy cognitive maps with neural networks for natural gas prediction. *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI* **2018-November**, 1026–1032 (12 2018). <https://doi.org/10.1109/ICTAI.2018.00158>
26. Poczęta, K., Papageorgiou, E.I.: Energy use forecasting with the use of a nested structure based on fuzzy cognitive maps and artificial neural networks. *Energies* 2022, Vol. 15, Page 7542 **15**, 7542 (10 2022). <https://doi.org/10.3390/EN15207542>

27. Potter, K., KARL, L.: Transfer learning and pretrained models (9 2023). <https://doi.org/10.31219/OSF.IO/9P7JV>
28. Redmon, J., Farhadi, A.: Yolo9000: Better, faster, stronger. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 **2017-January**, 6517–6525 (12 2016). <https://doi.org/10.1109/CVPR.2017.690>
29. Reynolds, D.A., Rose, R.C.: Robust text-independent speaker identification using gaussian mixture speaker models. IEEE Transactions on Speech and Audio Processing **3**, 72–83 (1995). <https://doi.org/10.1109/89.365379>
30. Sovatzidi, G., Vasilakakis, M.D., Iakovidis, D.K.: Automatic fuzzy graph construction for interpretable image classification. Proceedings - International Conference on Image Processing, ICIP pp. 3743–3747 (2022). <https://doi.org/10.1109/ICIP46576.2022.9897471>
31. Sovatzidi, G., Vasilakakis, M.D., Iakovidis, D.K.: Towards the interpretation of multi-label image classification using transformers and fuzzy cognitive maps. IEEE International Conference on Fuzzy Systems (2023). <https://doi.org/10.1109/FUZZ52849.2023.10309713>
32. Stempfle, T., Mangos, C., Feldmann, S., Kallien, L., Rössle, M., Jung, J.: Artificial intelligence-based error detection and adaption of shot curve parameters in the hot chamber high-pressure die casting process pp. 140–150 (9 2024). <https://doi.org/10.1109/ICRAS62427.2024.10654477>
33. Tan, M., Le, Q.V.: Efficientnet: Rethinking model scaling for convolutional neural networks. 36th International Conference on Machine Learning, ICML 2019 **2019-June**, 10691–10700 (5 2019)
34. Tziolas, T., Papageorgiou, K., Feleki, A., Theodosiou, T., Rapti, K., Papageorgiou, E., Pantoja, S., Cuinas, A.: Deep fuzzy cognitive maps for defect inspection in antenna assembly. Procedia Computer Science **232**, 97–106 (1 2024). <https://doi.org/10.1016/J.PROCS.2024.01.010>
35. Vasilakakis, M., Sovatzidi, G., Dimas, G., Iakovidis, D.K.: Towards the interpretation of convolutional neural networks for image classification using fuzzy sets. Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics **2022-October**, 2912–2917 (2022). <https://doi.org/10.1109/SMC53654.2022.9945092>
36. Weng, W., Zhu, X.: U-net: Convolutional networks for biomedical image segmentation. IEEE Access **9**, 16591–16603 (5 2015). <https://doi.org/10.1109/ACCESS.2021.3053408>