

Hydrocyclone Operational Condition Detection: Conceptual Prototype with Edge AI

Tomás Henrique Coelho e Silva^{1,3}^a, Ricardo Augusto Rabelo Oliveira²^b and Emerson Klippel¹^c

¹Vale SA, Parauapebas, Para, Brazil

²Computing Department, Federal University of Ouro Preto, Ouro Preto, Minas Gerais, Brazil

³PROFICAM, Instituto Tecnológico Vale, Brazil

Keywords: Convolutional Neural Networks, Deep Learning, Edge Device, Hydrocyclone, Transfer Learning, Vision Transformer.

Abstract: Hydrocyclones, vital in mineral processing plants, classify materials by size and density. Operational issues, like roping, can cause inefficiencies and financial losses. This paper explores computer vision techniques for the assessment of hydrocyclone underflow operational status. Testing revealed robust performance for both a Resnet-18 and a MobileViT-V2 model. An edge device was implemented for real-time inferences on a conceptual prototype that simulates underflow scenarios. The CNN models demonstrate high precision and recall, with an F1 Score over 92% for roping detection on the edge device. The research contributes to efficient hydrocyclone monitoring, addressing challenges in remote mining locations. The findings offer potential for further optimization and industrial implementation, enhancing processing plant reliability and mitigating financial risks associated with operational irregularities.

1 INTRODUCTION

Hydrocyclones are low-cost devices that are commonly used in mineral processing plants for material classification. They are used to select or classify material at a particular cut size and an optimum solids concentration percentage defined by downstream process requirements. The classification is achieved by opposing centrifugal and drag forces which move coarse and dense particles to the periphery, forcing them downwards to join the underflow and exit at the apex as part of the underflow. Meanwhile, fine and light particles as well as most of the water are directed to the upper exit at the vortex. (Luz et al., 2010)

The angle of the underflow discharge is an important diagnostic of the classification process condition in a hydrocyclone. In an ideal operation, an air core in the unit is generated and the apex discharge presents a fan shape, as shown in Figure 1 whereas under certain conditions the air core collapses and causes the underflow to be characterized by a rope shape, which indicates very high underflow density

and that coarse particles are being discharged with the overflow. (Napier-Munn and Centre, 1996).

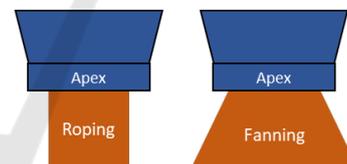


Figure 1: Hydrocyclones underflow states.

The issues associated with roping are derived from the higher classification cut size which may significantly reduce the efficiency of downstream processes and result in a lower metallurgical recovery. Another effect that may arise is the blocking of the spigot or of devices used in the overflow pipelines, which can lead to plant shutdowns (Concha et al., 1996). Both outcomes impair circuit performance and cause financial losses to the processing plant.

As a result, the monitoring of hydrocyclone operational status has been largely investigated for optimal performance, with a focus on a variety of techniques, which include ultrasound monitoring (Olson and Waterman, 2005), vibration (S. Mishra and Majumder, 2022), and image analysis (Janse van Vuuren et al., 2011). Even though some of these methods are com-

^a <https://orcid.org/0009-0007-4331-2281>

^b <https://orcid.org/0000-0001-5167-1523>

^c <https://orcid.org/0000-0003-0312-3615>

mercially available, their widespread adoption in the industry has been limited.

A recent promising approach has been developed by (Giglia and Aldrich, 2020) which involves the use of a convolutional neural network classifier with hydrocyclone underflow images. It accomplished high accuracy without requiring significant image preprocessing. However, the adoption of cloud-based processing in industrial settings for real-time monitoring and immediate action may encounter significant challenges, primarily revolving around two key points: latency and the dependency on continuous cloud connectivity. These issues arise from the time delay in transmitting data to and from centralized cloud servers and the necessity for uninterrupted cloud connectivity. These challenges become particularly pronounced in remote mining processing plants with limited or unreliable network connectivity.

Thus, this investigation extends the study by (Giglia and Aldrich, 2020) by testing computer vision algorithms on an edge device with no centralized processing. The objective is to make real-time inferences regarding a conceptual prototype that simulates the operational status of a hydrocyclone underflow. In addition, a hybrid model between CNNs and Vision Transformers (ViT), MobileViT-V2, was also evaluated on the test dataset.

The paper is structured as follows: the next section describes the theoretical reference. The procedures and research design are explained in section 3. The results of the experiments are presented and discussed in section 4. Section 5 includes the conclusions.

2 THEORETICAL REFERENCE

The mining industry, as illustrated by (Zhang et al., 2021), leverages computer vision algorithms for various applications such as materials classification, identification of asset failure, analysis of ore constituents, and so on. This section addresses the theoretical foundations of two prominent deep learning architectures employed in computer vision tasks — Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) — and explores the concept of edge AI.

2.1 Neural Network Models

Since the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), CNNs have become a cornerstone in vision-related tasks due to their robust performance demonstrated on this benchmark dataset, as shown by (Krizhevsky et al., 2012), and they are widely used in various computer vision applications.

CNNs are specifically designed to process data with a grid-like structure, such as images, by employing convolutional layers to extract hierarchical features from input data via kernel convolutions, exhibiting memory-efficient properties like parameter sharing and sparse connections (Goodfellow et al., 2016). Subsequent pooling layers merge similar features to reduce dimensions and enhance invariance to input distortions. Fully connected layers at the network's end utilize the extracted features for specific tasks like classification.

The Transformer, introduced by (Vaswani et al., 2017), revolutionized neural network architectures by relying solely on attention mechanisms for sequential data processing, enabling flexible dependency modeling without considering input distance and emerging as a state-of-the-art solution for Natural Language Processing (NLP). Vision Transformers, pioneered by (Dosovitskiy et al., 2021), extend the Transformer concept to image recognition by treating images as sequences of patches instead of image-specific architectural biases, surpassing many CNN-based image classification methods in accuracy and computational efficiency.

Further advancements include hybrid models that combine self-attention mechanisms with CNNs to capture both long-range dependencies and local information. Research comparing CNNs and Transformers for visual tasks found it challenging to declare a winner, but highlighted hybrid models for their efficacy and cost-effectiveness, leveraging strengths from both architectures while mitigating their limitations (Moutik et al., 2023).

Training CNNs and ViTs from scratch often requires a large image dataset. Transfer learning offers an effective strategy to overcome this limitation by adapting pre-trained models, originally trained on large datasets, to new settings. This approach involves substituting the classification segment of the pre-trained model with untrained layers tailored for the specific classification task in the new setting, thereby accelerating learning with limited data.

2.2 Edge Artificial Intelligence

Edge AI (EI) merges Edge computing (EC) and Artificial Intelligence (AI), processing AI computations on edge devices near data sources. EI shifts data processing tasks from the cloud to the edge of the network, relieving the cloud of the burdens of data processing, storage, and computing processes (Hua et al., 2023). The relationship between EC and AI benefits both domains as EC provides an ideal environment for AI by enhancing data availability and accommo-

dating diverse applications, while AI optimizes EC by processing multimodal data, detecting patterns, and improving decision-making (Singh et al., 2022).

Edge AI offers several advantages over cloud-based computing, including lower latency, enhanced privacy, cost-effectiveness, and improved reliability. Localized computation reduces latency and boosts privacy and security by processing sensitive data on-site, while eliminating constant data transmission to centralized data centers saves costs with communication and cloud infrastructure. It also enhances reliability, with the ability to operate independently without a continuous internet connection (Singh and Gill, 2023). However, limited processing and storage capacity in edge devices remains a challenge, prompting research into processor acceleration and model adaptation techniques to optimize AI frameworks for resource-constrained devices (Deng et al., 2020).

2.3 Related Work

Computer vision applications for failure and operational status detection are currently undergoing substantial expansion. Some of these developments are covered in this topic.

(Giglia and Aldrich, 2020) applied a CNN-based model to assess the operational status of a hydrocyclone based on its underflow images. They employed transfer learning to industrial underflow video frames to construct two-state classifiers (fanning or roping) using VGG-19 with a CNN classifier and ResNet50 with an SVM classifier. The CNN classifier attained accuracies of 76.0% and 82.8% on video footage datasets, while an ensemble model combining CNN and SVM classifiers achieved accuracies of 98.2% and 84.4%, highlighting the effectiveness of their approach. However, test results for roping with industrial images were not provided.

(Liu and Aldrich, 2023) applied Vision Transformers to classify froth flotation images associated with different operational regimes, demonstrating competitive accuracy compared to CNN-based approaches. Additionally, (Hütten et al., 2022) conducted a comprehensive comparison between CNNs and Vision Transformers for industrial visual inspection tasks, concluding that Vision Transformers outperformed CNNs, demonstrating no significant difference in convergence speed and showcasing their efficacy in handling small datasets.

Other applications have recently been developed for defect detection and classification by employing CNN-based, ViT-based, or hybrid architectures. Examples include the classification of maize seeds, as explored by (Chen et al., 2022) using ViT, strip steel

surface defect classification by (Li et al., 2022) utilizing hybrid models, and PCB defect detection by (An and Zhang, 2022) introducing a ViT-based model achieving state-of-the-art results. Moreover, Edge AI applications, like (Klippel et al., 2022) implementing a CNN-based model for conveyor belt rip detection and (Li et al., 2023) proposing a MobileViT-based architecture for real-time plant disease detection, highlight the diverse and expanding range of industries benefitting from advanced computer vision models.

This article approaches an important problem in the mining industry, hydrocyclone operational status detection, and proposes a solution that leverages Edge AI to overcome potential latency and connectivity issues in remote plants while harnessing the robust generalization capacity of Artificial Intelligence algorithms.

3 METHODOLOGY

This section provides an in-depth exploration of the procedures and research design. Firstly, it discusses the selected edge device for the experiments. It then outlines the development of a conceptual prototype designed to assess the feasibility of the proposed approach. Finally, the section explores the employed framework for model training and deployment.

3.1 Edge Device

The edge device chosen for the experiments is the Sipeed Maix Dock II, a cost-effective board designed for AI applications. It is powered by the Allwinner V831 chip, operates on Linux, and features a single-core ARM Cortex-A7 with 64MB DDR2 RAM, supporting speeds of up to 800 MHz. It incorporates a Neural Processing Unit (NPU) dedicated to executing AI tasks at a performance level of up to 0.2 TOPS. The device is equipped with various peripherals, including an analog microphone, a 3-axis acceleration sensor, and a 2MP HD camera, which enhances its versatility for diverse applications.

3.2 Conceptual Prototype

A conceptual prototype was developed to assess the feasibility of the proposed approach and to serve as a proof of concept for potential industrial applications. The prototype simulated a hydrocyclone underflow by using a hose with an adjustable sprayer placed 90 cm from the camera. A black panel was positioned behind the hose to improve contrast. The flow opening

of the hose was manipulated to emulate both roping and fanning phenomena observed in a hydrocyclone.

For image capture, two distinct devices were employed. The Samsung S22 smartphone camera, capable of recording high-definition video at 1920 x 1080 pixels resolution and a frame rate of 30 frames per second (fps), served to capture dynamic footage of the simulated underflow. Additionally, to provide a more granular and varied dataset for in-depth analysis, the Maix Dock II edge device was also configured to capture images at a fixed interval of 0.2 seconds.

Furthermore, the spatial relationship between the hose and the camera was systematically adjusted. Horizontal and vertical variations were explored, along with changes in the distance between them within a range of 15 cm. This experimental design, shown in Figure 2, aimed to comprehensively simulate diverse conditions resembling those encountered in real hydrocyclone underflow scenarios.

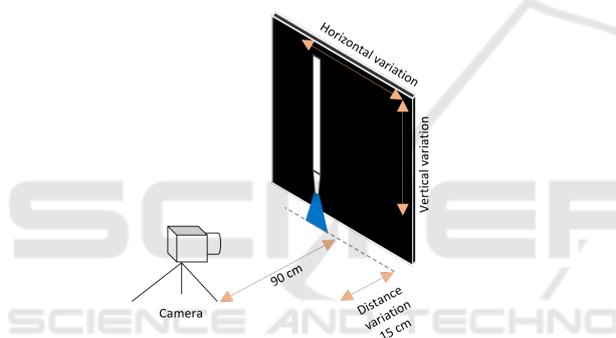


Figure 2: Diagram of the conceptual prototype.

3.3 Training and Deployment Framework

The training framework was implemented in Python. This subsection provides an in-depth discussion of the key components and processes within the framework.

For the construction of the inference models, two different architectures were chosen: ResNet-18 and MobileViT-V2. The Convolution Neural Network selected was ResNet-18, introduced by (He et al., 2015). It is an 18-layer network that adds residual learning blocks into ConvNets. This design enhances accuracy by addressing challenges related to increased network depth, thereby preventing accuracy saturation that might occur as the depth of the network increases. The hybrid Visual Transformer model was MobileViT-V2, introduced by (Mehta and Rastegari, 2022). It combines the strengths of CNNs and Vision Transformers (ViTs) to capture spatial hierarchies and self-attention mechanisms, respectively, offering ver-

satility in image processing.

The initial phase of the framework involves capturing images from the video footage made with the smartphone camera by utilizing Python's Decord library and its VideoReader module. Images are captured at intervals of 10 frames and, employing OpenCV-Python, are subsequently exported into the computer.

Images from both image capture devices are organized into three folders corresponding to their class: background, roping, and fanning. They are then loaded and divided into training and validation datasets using Torchvision's ImageFolder dataset generator in an 80-20 ratio. The images are labeled according to their class. The training dataset undergoes transformation and augmentation, including resizing to 224 x 224, random horizontal flipping, random rotations within a 90-degree range, and random adjustments to brightness, contrast, saturation, and hue. These transformations are applied using Torchvision's Transforms V2 and ColorJitter V2 transforms. Finally, normalization is performed using ImageNet's mean and standard deviation statistics. The validation dataset undergoes resizing and normalization only.

Another video footage is utilized to generate images for the testing dataset, and these images undergo the same transformations as the validation set. Subsequently, the testing dataset is employed for an unbiased evaluation of the models' performance on new and unseen data. This approach facilitates a proper evaluation of the model, even without utilizing K-fold validation, thereby saving computational costs.

As a result, the complete dataset comprises a total of 1730 images, with 740 images each for roping and fanning, and 250 images for the background. They are randomly split into training and validation sets. Additionally, there are 260 images in the testing dataset, with 100 for roping, 100 for fanning, and 60 for the background class.

Both models, initially pre-trained on the ImageNet dataset, were adapted using transfer learning through two distinct approaches. In the first method, termed as partially retrained, only the parameters of the final layer, the classification layer, were updated using the training dataset, while the rest of the model parameters remained fixed. Conversely, the fully retrained approach involved updating all parameters of the models using the training data. Both training processes employed a low learning rate of 0.001, with an exponential decay factor of 0.977 applied to it. Additionally, Stochastic Gradient Descent optimizer and cross-entropy loss were utilized during training.

The ResNet-18 encompasses a total of 11,178,051

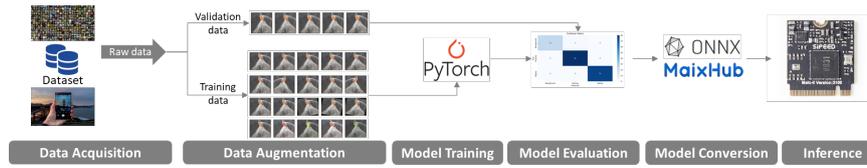


Figure 3: Training and deployment framework.

parameters. Similarly, the MobileViT-V2 model consists of a total of 4,390,380 parameters. Their classification layer is composed of 1,539 parameters.

Three metrics are used to evaluate the models: precision, recall, and F1 Score. Precision assesses the model's ability to identify instances of a particular class correctly while Recall evaluates the model's ability to identify all instances belonging to that class. The F1 score is a balanced combination of precision and recall. These metrics are calculated using equations 1, 2, and 3, respectively, taking into account true positives (TP), false positives (FP), and false negatives (FN) for each class. Weight averaging is used to aggregate each metric result across all classes, ensuring consideration of class imbalance.

$$Precision = \frac{TP}{(TP + FP)} \quad (1)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (2)$$

$$F1score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \quad (3)$$

Subsequently, the models are converted into the Open Neural Network eXchange (ONNX) format utilizing the 'torch.onnx.export' ONNX exporter in TorchScript. For inference on the Maix Dock II, they undergo an additional conversion from ONNX to AWNN format using the MaixHub tool. This final step ensured compatibility and optimized performance for deployment on the Maix Dock II.

The inference process occurs on the edge device through exposure to simulated scenarios of roping, fanning, and background. The device is set up to record images at a 0.2-second interval onto an SD memory, facilitating the assessment of the inference performance. A full illustration of the training and deployment framework can be found in Figure 3.

The ViT model was not implemented on the edge device because Maix Dock II does not support it. More efficient compression techniques are required to enable their implementation as has been proposed by (Song et al., 2022).

4 RESULTS

This section presents the results of the models for classifying the underflow status under three classes: background, roping, or fanning. The presentation is organized into two subsections for clarity and comparison. The first subsection details the results achieved on the test dataset, while the subsequent one outlines the outcomes obtained on the edge device.

4.1 Results on the Testing Dataset

The ResNet-18 partially retrained model was trained for 24 epochs. The training process was halted when no improvements in validation cross-entropy loss were observed for five consecutive epochs. Parameters from the model exhibiting the lowest loss during validation were chosen for the final model. Figures 4 and 5 show the corresponding losses and accuracies for both the training and validation phases. The model reached validation accuracies over 90% after only two epochs and reached losses close to 0.1.

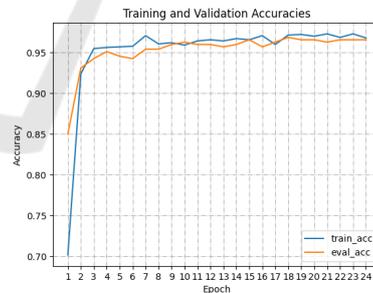


Figure 4: CNN partially trained model accuracy.

The confusion matrix for the test dataset is shown in Figure 6.

The ResNet-18 fully retrained model was trained for 15 epochs and the parameters from the model exhibiting the lowest loss during validation were chosen for the final model. Figures 7 and 8 show the corresponding losses and accuracies for both the training and validation phases.

After two epochs, the training reached accuracies over 98% and losses significantly lower than 0.1, which is even better than the partially trained model.

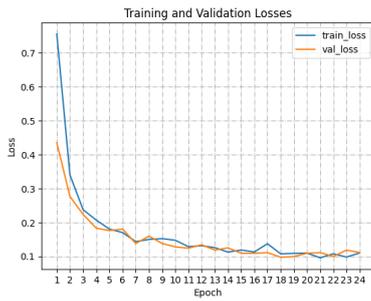


Figure 5: CNN partially trained model cross-entropy loss.

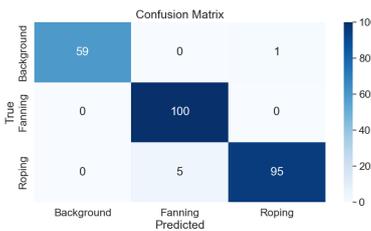


Figure 6: CNN part trained model confusion matrix - Test.

The confusion matrix for the test dataset is shown in Figure 9. It indicates robust performance and successful generalization, as only one sample from the test dataset was not correctly classified.

The MobileViT-V2 fully trained model was trained for 15 epochs. Parameters from the model exhibiting the highest accuracy during validation were chosen for the final model. The confusion matrix for the test dataset is presented in Figure 10. The model correctly assigned every sample from the test dataset, achieving 100 % accuracy.

Table 1 summarises the performance metrics results. The results suggest a robust performance of the fully trained models in comparison to the partially trained model. The MobileViT performed better than the CNN-based model but by a minimum margin, as it correctly predicted every image but the latter only had one sample mispredicted. It may be an indication that the MobileViT-V2 model can be a superior choice for this particular application.

Table 1: Models performance metrics on the testing dataset.

Model	Precision	Recall	F1 Score
CNN part trained	0.978	0.977	0.977
CNN fully trained	0.996	0.996	0.996
ViT fully trained	1.000	1.000	1.000

4.2 Results on the Edge Device

The Maix Dock II edge device with each of the CNN-based models was exposed to the conceptual prototype, which simulates hydrocyclone underflow condi-

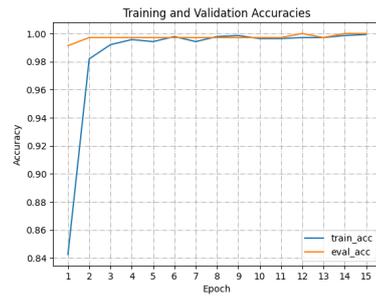


Figure 7: CNN fully trained model accuracy.

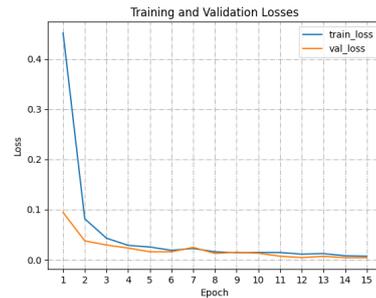


Figure 8: CNN fully trained model cross-entropy loss.

tions. The edge device was assessed on its ability to classify each class correctly.

The confusion matrix for the edge device results of the partially trained model is shown in Figure 11, while the results for the fully trained model are shown in Figure 12. Table 2 summarises the performance metrics results.

Table 2: Models performance metrics at the edge.

Model	Precision	Recall	F1 Score
CNN part trained	0.892	0.855	0.855
CNN fully trained	0.943	0.941	0.941

Both models exhibited decreased performance compared to their performance on the testing dataset, which is a common occurrence when quantized models are used, as they operate with lower precision. However, the results are promising, which is indicated by all performance metrics over 94% achieved by the fully trained model.

Furthermore, for the detection of roping, which is the critical state that the device needs to be able to detect, both partially and fully trained models exhibited F1 scores above 92%, as summarized in Table 3. This is an encouraging result for further improvements towards implementation in industrial settings.

In summary, all three models demonstrated robust performance on the testing dataset, with an F1 score exceeding 97%, indicating their ability to accurately identify each classes. The fully-trained models outperformed the partially-trained one, benefiting from

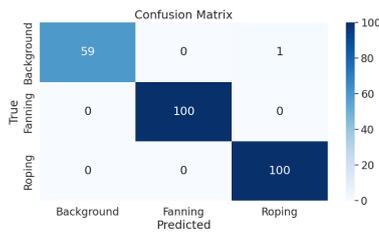


Figure 9: CNN fully trained model confusion matrix - Test.

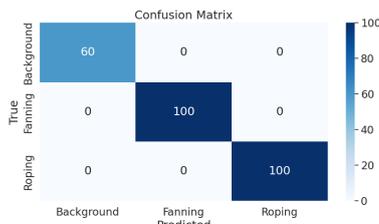


Figure 10: MobileViT-V2 model confusion matrix - Test.



Figure 11: CNN part trained model confusion matrix-Edge.



Figure 12: CNN fully trained model confusion matrix-Edge.

comprehensive optimization across all parameters.

Despite slightly reduced performance on the edge, the fully trained CNN model achieved over 94% accuracy overall and above 92% for roping detection, indicating encouraging results for developing models trained on industrial hydrocyclone images for real-world deployment. Further hyperparameter tuning, such as the number of epochs and learning rate, could lower cross-entropy losses and improve performance when deployed on the edge device. Exploring the use of ViT models on the edge is worth considering given their excellent performance on the test dataset.

However, for industrial deployment, some practical considerations need to be considered. First, the device would require an unobstructed view of the hy-

Table 3: Models performance metrics for roping.

Model	Precision	Recall	F1 Score
CNN part trained	0.889	0.977	0.931
CNN fully trained	0.922	0.932	0.927

drocyclone apex for effective operation. Additionally, lighting, hydrocyclone sizes and design, and different classified materials should be further explored when training and deploying this concept.

5 CONCLUSIONS

In conclusion, this paper has explored the critical role of hydrocyclones in mineral processing plants and the potential issues associated with roping, which can lead to reduced efficiency and financial losses. The study presented an approach to monitor hydrocyclone operational status using edge computing and computer vision techniques.

The investigation tested the use of two neural network models, ResNet-18 and MobileViT-V2, which were examined for their effectiveness regarding the operational status detection of a hydrocyclone. The Resnet-18 model was also implemented on an edge device, specifically the Sipeed Maix Dock II, and tested on a conceptual prototype that simulates the behavior of the underflow of a hydrocyclone. It revealed a slight degradation in accuracy, likely attributed to quantization effects, but the overall findings support the feasibility of deploying these models in real-world scenarios, as meaningful results, F1 scores over 94% overall and 92.7% for roping detection, were obtained.

This research serves as a pilot for developing solutions to optimize mineral processing plant performance and address challenges in remote locations. Future work should focus on training the models with a dataset comprising images of the underflow from various industrial hydrocyclones. This dataset should encompass a diverse range of lighting conditions, underflow rates, camera angles, and positions relative to the apex, as well as various backgrounds to effectively enhance the model's generalization capacity across different scenarios. Additionally, refining the model training process to achieve even lower validation losses could mitigate accuracy degradation when the models are quantized and deployed on the edge device. Furthermore, evaluating the performance of Vision Transformers (ViT) could also lead to significant improvements on the Edge device.

Finally, conducting rigorous testing on an industrial-scale hydrocyclone operating under varying conditions is essential to validate the effectiveness

and applicability of the proposed solutions in real-world operational scenarios.

ACKNOWLEDGEMENTS

The authors would like to thank FAPEMIG, CAPES, CNPq, Instituto Tecnológico Vale, and the Federal University of Ouro Preto for supporting this work. This study was partially funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) finance code 308219/2020-1.

REFERENCES

- An, K. and Zhang, Y. (2022). Lpvit: A transformer based model for pcb image classification and defect detection. *IEEE Access*, 10:42542–42553.
- Chen, J., Luo, T., Wu, J., Wang, Z., and Zhang, H. (2022). A vision transformer network seedvit for classification of maize seeds. *Journal of Food Process Engineering*, 45(5):e13998.
- Concha, F., Barrientos, A., Montero, J., and Sampaio, R. (1996). Air core and roping in hydrocyclones. *International Journal of Mineral Processing*, 44-45:743–749. Comminution.
- Deng, S., Zhao, H., Fang, W., Yin, J., Dustdar, S., and Zomaya, A. Y. (2020). Edge intelligence: The confluence of edge computing and artificial intelligence. *IEEE Internet of Things Journal*, 7(8):7457–7469.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale.
- Giglia, K. and Aldrich, C. (2020). Operational state detection in hydrocyclones with convolutional neural networks and transfer learning. *Minerals Engineering*, 149:106211.
- Goodfellow, I. J., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge, MA, USA. <http://www.deeplearningbook.org>.
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition.
- Hütten, N., Meyes, R., and Meisen, T. (2022). Vision transformer in industrial visual inspection. *Applied Sciences*, 12(23).
- Hua, H., Li, Y., Wang, T., Dong, N., Li, W., and Cao, J. (2023). Edge computing with artificial intelligence: A machine learning perspective. *ACM Comput. Surv.*, 55(9).
- Janse van Vuuren, M., Aldrich, C., and Auret, L. (2011). Detecting changes in the operational states of hydrocyclones. *Minerals Engineering*, 24(14):1532–1544.
- Klippel, E., Oliveira, R. A. R., Maslov, D., Bianchi, A. G. C., Delabrida, S. E., and Garrocho, C. T. B. (2022). Embedded edge artificial intelligence for longitudinal rip detection in conveyor belt applied at the industrial mining environment. *SN Computer Science*, 3(4):280.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.
- Li, G., Wang, Y., Zhao, Q., Yuan, P., and Chang, B. (2023). Pmvt: a lightweight vision transformer for plant disease identification on mobile devices. *Frontiers in Plant Science*, 14:1256773.
- Li, S., Wu, C., and Xiong, N. (2022). Hybrid architecture based on cnn and transformer for strip steel surface defect classification. *Electronics*, 11(8):1200.
- Liu, X. and Aldrich, C. (2023). Flotation froth image recognition using vision transformers. *IFAC-PapersOnLine*, 56(2):2329–2334. 22nd IFAC World Congress.
- Luz, A. B. d., Sampaio, J. A., and França, S. C. A. (2010). Tratamento de minérios.
- Mehta, S. and Rastegari, M. (2022). Separable self-attention for mobile vision transformers.
- Moutik, O., Sekkat, H., Tigani, S., Chehri, A., Saadane, R., Tchakoucht, T. A., and Paul, A. (2023). Convolutional neural networks or vision transformers: Who will win the race for action recognitions in visual data? *Sensors*, 23(2).
- Napier-Munn, T. and Centre, J. K. M. R. (1996). *Mineral Comminution Circuits: Their Operation and Optimisation*. JKMRRC monograph series in mining and mineral processing. Julius Kruttschnitt Mineral Research Centre.
- Olson, T. and Waterman, R. (2005). Hydrocyclone roping detector and method. United States Patent US6983850B2.
- S. Mishra, M. H. T. and Majumder, A. K. (2022). Development of a vibration sensor-based tool for on-line detection of roping in small-diameter hydrocyclones. *Mineral Processing and Extractive Metallurgy Review*, 0(0):1–20.
- Singh, A., Saini, K., Nagar, V., Aseri, V., Sankhla, M. S., Pandit, P. P., and Chopade, R. L. (2022). Chapter sixteen - artificial intelligence in edge devices. In Raj, P., Saini, K., and Surianarayanan, C., editors, *Edge/Fog Computing Paradigm: The Concept Platforms and Applications*, volume 127 of *Advances in Computers*, pages 437–484. Elsevier.
- Singh, R. and Gill, S. S. (2023). Edge ai: A survey. *Internet of Things and Cyber-Physical Systems*, 3:71–92.
- Song, H., Wang, Y., Wang, M., and Wang, Z. (2022). Ucvit: Hardware-friendly vision transformer via unified compression. In *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2022–2026.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- Zhang, M., Shi, H., Zhang, Y., Yu, Y., and Zhou, M. (2021). Deep learning-based damage detection of mining conveyor belt. *Measurement*, 175:109130.