

Article

# Centrifugal Pump Fault Detection with Convolutional Neural Network Transfer Learning

Cem Ekin Sunal <sup>1</sup>, Vladan Velisavljevic <sup>1</sup>, Vladimir Dyo <sup>2,\*</sup>, Barry Newton <sup>3</sup> and Jake Newton <sup>3</sup>

<sup>1</sup> School of Computer Science and Technology, University of Bedfordshire, Luton LU1 3JU, UK; vladan.velisavljevic@beds.ac.uk (V.V.)

<sup>2</sup> Department of Electronic Engineering, Royal Holloway, University of London, Egham TW20 0EX, UK

<sup>3</sup> Uptime Systems Ltd., Leighton Buzzard LU7 4WG, UK; barry.newton@uptime.uk.com (B.N.)

\* Correspondence: vladimir.dyo@rhul.ac.uk

**Abstract:** The centrifugal pump is the workhorse of many industrial and domestic applications, such as water supply, wastewater treatment and heating. While modern pumps are reliable, their unexpected failures may jeopardise safety or lead to significant financial losses. Consequently, there is a strong demand for early fault diagnosis, detection and predictive monitoring systems. Most prior work on machine learning-based centrifugal pump fault detection is based on either synthetic data, simulations or data from test rigs in controlled laboratory conditions. In this research, we attempted to detect centrifugal pump faults using data collected from real operational pumps deployed in various places in collaboration with a specialist pump engineering company. The detection was done by the binary classification of visual features of DQ/Concordia patterns with residual networks. Besides using a real dataset, this study employed transfer learning from the image detection domain to systematically solve a real-life problem in the engineering domain. By feeding DQ image data into a popular and high-performance residual network (e.g., ResNet-34), the proposed approach achieved up to 85.51% classification accuracy.

**Keywords:** machine learning; Internet of things; centrifugal pump; condition monitoring



**Citation:** Sunal, C.E.; Velisavljevic, V.; Dyo, V.; Newton, B.; Newton, J. Centrifugal Pump Fault Detection with Convolutional Neural Network Transfer Learning. *Sensors* **2024**, *24*, 2442. <https://doi.org/10.3390/s24082442>

Academic Editors: Hiroharu Kawanaka, Chan-Yun Yang and Hooman Samani

Received: 6 March 2024

Revised: 7 April 2024

Accepted: 9 April 2024

Published: 11 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Centrifugal pumps are ubiquitous in industrial and domestic applications and represented a 51% market share of the USD 98.02 billion pump market in 2020 [1]. As with any other mechanical equipment, centrifugal pumps are subject to wear and can develop faults related to bearing, sealing and blockage, in addition to electrical faults within the driving motor. The timely detection of these faults is essential to prevent accidents and disruptions in industrial and domestic processes.

Motor current signature analysis (MCSA) is an established condition-monitoring method based on monitoring distinct electric current patterns of the stator in the induction motor [2], which drives the pump, using a combination of signal processing, statistical techniques and machine learning [3]. The MCSA-based systems can be deployed by simply attaching current clamps to power supply wires, which can be done away from the pump itself. Compared with other sensing modalities, such as vibration, acoustic [4] and pressure, MCSA does not require physical access to the pump, which makes it practical and sometimes the only feasible method for some applications, such as submerged sewage pumps. MCSA has been applied for the detection of broken rotor bars [5–7], stator winding faults [8], cavitation faults [9,10], blockages [10–12], bearing faults [13], impeller defects [14], shaft wear [15], and inner and outer race faults [16].

Recently, the emergence of advanced machine learning algorithms, in combination with low-cost sensing hardware, has opened up new opportunities for intelligent preventative maintenance. However, many machine learning-based MCSA fault detection works generally relied on simulation or synthetic data [17]. The studies that used MCSA on real

physical motor/pump data [17] utilised test rigs in controlled laboratory conditions. Often, the faults would be artificially generated by drilling a hole in a rotor bar to mimic a broken rotor bar condition [18], using polystyrene to clog impellers [19], etc., The prior work on applying machine learning to real pumps that have been deployed in real-world conditions is limited and not well documented. This is often due to the scarcity of pump-related datasets, cost and significant practical difficulties in collecting data from real pumps [17].

The presented study applied a machine learning algorithm to reliably detect motor faults using data obtained from several operational pumps running at various locations in the UK. This paper first presents a distributed IoT system designed to sense, transmit, store and process MCSA data using low-cost hardware developed in this project. It then describes a unique dataset collected from centrifugal pumps deployed in the UK over a period of almost 2 years between 2021 and 2023. The dataset included data from non-faulty operational pumps; pumps with diagnosed faults; and intentionally damaged pumps, including damage such as impeller damage, stator winding faults, excessive vibrations, cavitations and/or bearing damage faults. A novel fault detection method is then proposed based on DQ pattern analysis that built upon work by [11,20–22]. The method extracts DQ patterns and feeds them into convolutional neural network transfer learning for analysis to provide the binary classification of different type of faults. The paper makes the following contributions:

- A novel IoT-based system for measuring, collecting and storing MCSA data from operational pumps.
- A novel machine learning method for pump fault detection that uses convolutional neural network transfer learning on DQ patterns.
- An evaluation of the proposed algorithm using custom dataset from operational pumps.
- An experience in practical challenges and limitations in implementing such a system.

To the best of our knowledge, this is the first study that applied machine learning for fault detection using real datasets from operational pumps.

## 2. Related Work

### 2.1. IoT-Based Induction Motor Monitoring

The application of machine learning for the condition monitoring of induction motors is an active research area. A comprehensive analysis and review of machine learning for fault detection in centrifugal pumps is available in [17]. Considering that an inductive motor is an essential component that drives a centrifugal pump, the analysis below includes relevant work on MCSA-based induction motor fault detection. The emergence of advanced machine learning algorithms, in combination with low-cost sensing hardware, has opened up new opportunities for intelligent preventative maintenance. Most prior work on IoT-based induction motor monitoring focuses on operational conditions, electrical efficiency or output power of the motors. With few exceptions, most experimental evaluation has been done in a laboratory test bench. Sen et al. [23] designed a system for induction motor monitoring with current measurements. Similarly to our work, the authors evaluated their system on deployed operational systems, which, in their case, was in a textile factory. However, the work focussed on hardware and software design but did not provide much details about the algorithms or dataset. Cano-Ortego and Sanchez-Sutil [24] designed a low-cost system to acquire the voltage, current and mechanical parameters for improving the energy efficiency and power factor of induction motors. The system protects the motor against overloads and uses LoRa for long range communication and stores data in Google Cloud. The presented prototype was used in a case study using electrical machines in the university laboratory. Carvalho et al. [25] proposed a method for monitoring the efficiency and output power of three-phase induction motors using three-phase voltages, currents and the equivalent circuit of the machine. The system was evaluated on an experimental test bench. Sridhar et al. [26] describes a system for fault classification of induction motors by processing stator current measurements using neural networks.

Kunthong et al. [27] implemented an IoT system for monitoring induction motor vibration, current and temperature in electric vehicles.

## 2.2. DQ Pattern Analysis

The presented approach was inspired by the discussion of DQ patterns and related motor conditions by Irfan et al. in [11,20,21], as well as a Park vector application for DQ [22]. Irfan et al. [11] used a three-phased 1.5 hp, 3450 rpm and 60 Hz centrifugal pump and proposed an electric diagnostic technique to detect CP faults without needing extra sensors. They measured the three-phase line currents with the voltage; then, they transformed the data to two-phase DQ patterns. The authors showed that the shapes of the obtained DQ pattern plots were affected by the motor faults, for example, the impeller or blockage faults disturbed the hexagonal DQ plot to a fan-shaped circle. They obtained their dataset of 1000 samples, which were sampled at a 4000 Hz rate by using a PXIe-1082 data acquisition module [11]. Their fault detection attempt was undertaken with pattern classification based on statistical indices after plotting the DQ patterns.

Therefore, it is believed that DQ is an intuitive method for detecting induction motor centrifugal pump faults. Thanks to their distinct characteristics discussed in Section 3.2, it is thought that a smaller-scale state-of-the-art convolutional neural network (CNN) that does not require a high amount of resources and can be deployed to consumer-grade hardware will be satisfactory to solve the detection problem. As CNNs are popular tools for computer vision problems thanks to its computational efficiency and parameter-sharing capability, we think they can be scaled to consumer-grade devices usage to detect simple image pattern recognition problem posed by DQ patterns. Given that their features are not as complex as the real world objects that ImageNet has, we believe that relatively small but proven state-of-the-art residual networks, like ResNet-34, can be used under transfer learning to efficiently solve the problem. Residual networks are the improved version of CNNs, where they use residual layers to keep gradients from vanishing [28] and allow for the development of deeper-layered CNNs.

In order to detect the motor faults in the induction motors or centrifugal pumps, the usage of CNNs were explored by researchers like Valtierra-Rodriguez et al. [7]. They used a CNN to detect broken rotor bars with a short-time Fourier transform signals map of the current signatures, with a final accuracy as high as 100%. They used a WEG-00136APE48T in-test motor that had two poles, 28 bars and a nominal power of 1 hp, and worked with 220 VAC at 60 Hz. Their dataset, however, only had 400 current signals, with a 3:1 training-to-validation dataset ratio. For more detailed analysis of using images for classification for industrial, health and other applications, we refer the reader to other work, e.g., [29–33].

## 3. Methods

This section presents the methods used in the creation of the model's pipeline. We first describe the Park transformation, which was used to convert three-phase measurements into two components named D and Q. The visualisation of these two components on a two-dimensional plot produces a circular shape for a healthy motor, whereas any distortions may indicate a faulty condition. The proposed approach uses a convolutional neural network algorithm to capture these distortions to detect faults. It should be noted that the idea of using images for classification purposes have been actively explored in various other domains, e.g., [30,31].

### 3.1. Park Transformation

For the data transformation we decided to employ Park transform. The transform can be used by converting the three-phase motor current signature to a two-phase system that has two components named D and Q in order to describe three-phase IM phenomena with Park's vector [6]:

$$x_d(t) = \sqrt{\frac{2}{3}}x_a(t) - \frac{1}{\sqrt{6}}(x_b(t) - x_c(t)) \quad (1)$$

$$x_q(t) = \frac{1}{\sqrt{2}}(x_b(t) - x_c(t)) \quad (2)$$

The space vector:

$$x_s(t) = \sqrt{x_d^2(t) + x_q^2(t)} \quad (3)$$

where  $x_d$  and  $x_q$  are Park's vector components. The component of these vectors are derived from the three weighted phases and their subtractions from each other [6]. The Park transform is essentially a conversion of three-phase motor current data to two components called  $d$  and  $q$ . DQ/Concordia patterns are the plots of  $d$  and  $q$  against each other. They have various (disturbed) shapes that give us an indication whether the sampled motor signature has a failure or not. The components D and Q are then used to generate a 2D image, which is used for fault classification, as described in the following section.

### 3.2. The Identification of DQ Patterns

DQ has several implementation areas in the literature and the ability to detect several motor faults. Given that it can be applied to the three-phase induction motors that any system (e.g., wind turbines) can have, it has the potential for a wide range of usage. Each fault can have distinct DQ patterns. Currently there are only a handful of publications that demonstrate the change in DQ pattern between healthy and faulty, and in this section, they will be discussed. According to the authors of [11,20], the shape of a healthy motor's DQ data is perfectly circular. Aside from the ability to detect the shape of healthy motors, a DQ pattern is reported to demonstrate the presence of impeller, blockage, BRB and short-circuit turn failures.

BRBs and/or short circuit between turns faults: For every turn, the tracked MCSA plot produces a slightly altered version of every cycle and causes an elliptical shape of the Park transform [21]. The difference can also be seen more clearly between healthy and a BRB fault problem in [20]. The DQ plot does not overlap the area it has already passed due to fault-induced phases.

Impeller faults and pipe blockages: The authors of [11] used DQ plots to illustrate how faults gradually change the shape of the DQ plots. They used a (hand-valve) half-pipe block, full-pipe block, (artificially) damaged impeller surface and 120-degree angled damaged impeller surface. According to their observations in line plots, the healthy shape of a pump is relatively hexagonal. However, this status changes when the aforementioned faults are introduced. They observed that for every increasingly severe fault iteration, the hexagonal shape is lost and a fan shape is constructed. As discussed earlier, this newly formed fan shape is not perfect, and like BRB faults, for every period, the plot does have a slightly different radius and period length. This causes the fan shape to be disrupted and not uniform in parallel to the increasing fault severity.

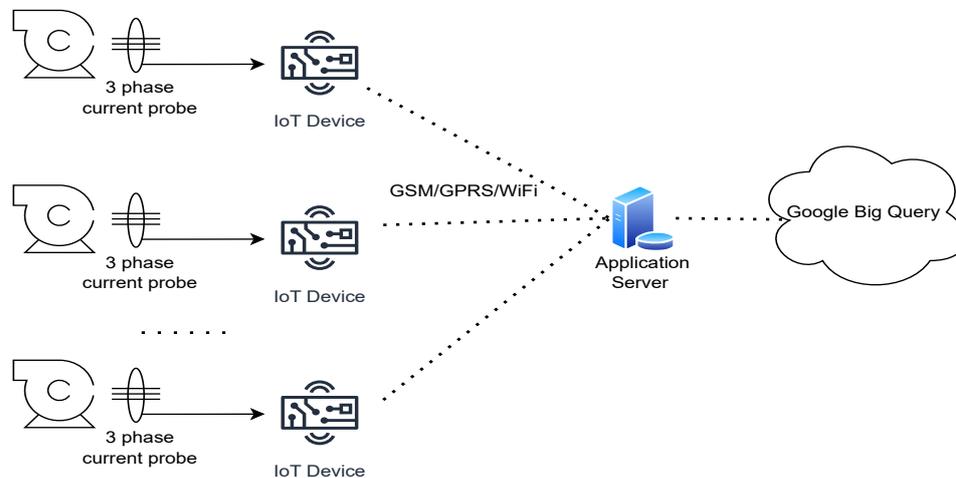
Therefore, our aim was to not only focus on these faults, but also investigate the affects of other faults on the DQ plot. In the end, we strived to create a fault-blind model that can ideally detect any fault without prior knowledge or at least with limited knowledge.

## 4. Implementation

### 4.1. System Architecture

The system consists of custom-developed IoT devices developed by Uptime Systems [34] and deployed at five customer sites, a Google Big Query Database and an application server, as shown in Figure 1. The IoT devices were Arduino-based and used SCT013 20A/1V transducers for AC current sensing. A number of devices were fabricated, deployed and maintained over an almost 2-year period from 2021 to 2023. The IoT devices were deployed near the pump control panels, measured the AC current in three phases and then transmitted the raw data to the application server over a GSM/GPRS link. The measurements were performed on each pump activation, at most every 60 min at 1500 Hz, 3000 Hz and 4500 Hz sampling rates, collecting 1500 samples for each phase per measure-

ment. These sampling rates were consistent with prior work, which ranged from 1500 Hz to 50,000 Hz for detecting various faults, such as broken rotor bar, inter turn and bearing faults. As some devices were deployed in basements or geographical areas with no or poor cellular connectivity, some devices were equipped with Wi-Fi modules to connect to a local WLAN.



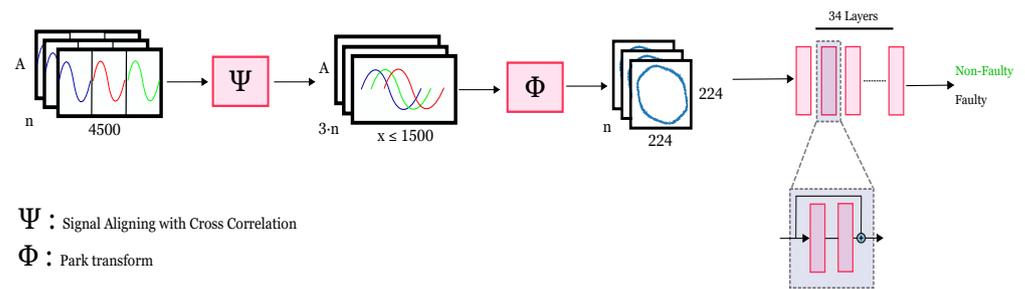
**Figure 1.** System architecture.

Finally, a custom application server was developed in the Flask framework to process incoming data streams; generate alarms; store/retrieve sensor data from the Google Big Query Database; and finally, visualise the data and system status on a dashboard. The description and characteristics of the pumps, as well as their faults, are provided later in the following subsections.

#### 4.2. Feature Extraction and Classification

The detection process starts with the data retrieval from the database, as shown in the left-most block in Figure 2. Each batch had a dimension of  $A \cdot 4500 \cdot n$ , where  $A$ , 4500 and  $n$  mean the amplitude, concatenated three phase data points and the total number of collected signatures, respectively. Given that the collected data for each phase were not aligned in phase due to the hardware limitations, we developed a function  $\psi$  to align the three phases. The original data shape had a dimension of  $1500 \cdot 3$ , where 1500 was the total amount of data points sampled per phase and 3 was the total phase number. The aligning process ( $\psi$ ) was done via the calculation of the cross-correlation of each signal with respect to each other. Then, the optimization step to find the minimum amount required to shift and cut the signal from their ends to align was performed. All sequences were calculated to minimize the data loss due to alignment (e.g., aligning all phases with respect to the first, second or third phase). The obtained signals were aligned with the dimensions of  $A \cdot (x \leq 1500) \cdot (3 \cdot n)$ , where  $x$  was the total amount of data points per phase. As phase alignment can lead to data loss, the actual number of data point for each phase could be less than 1500. The choice of 1500 samples per phase was due to hardware limitations and can be increased in the future.

Next, the DQ components were extracted from the aligned three-phase data using the Park transform (marked as  $\phi$  in Figure 2), resulting in batches of  $(x \leq 1500) \cdot (2 \cdot n)$  dimensions, with a  $d$  and  $q$  component for every point. The D and Q components were then plotted as a gridless RGB image for further pre-processing by a machine learning model. The pre-processing consisted of resizing, center cropping to an appropriate model input dimension ( $224 \cdot 224 \cdot 3$ ) and normalisation, before finally being fed to the model (e.g., ResNet-34).



**Figure 2.** Machine learning pipeline.

The final step was the ML detection through feeding DQ images through consumer-grade hardware and a state-of-the-art residual network. Although the current model results were classified as either “faulty” or “non-faulty”, the final layer can be extended to hold many fault classes, as mentioned in the future work section. One particular benefit of plotting the DQ values and feeding them as images is that the plot will not have axis values. This means that the pumps with higher voltages will not cause any overfitting with their higher DQ values. This can help to eliminate the bias based on the largeness and purely focus on the image features.

#### 4.3. Data Processing and Augmentation

The classification labels were set to “faulty/1” and “non-faulty/0”. The training-to-validation ratio was set to  $\sim 3:1$  and the faulty-to-non-faulty ratio was set to  $\sim 1:1$ . The validation and training datasets were prepared with 1500 and 3000 Hz sampling frequencies. The testing dataset consisted of two pumps that were sampled with three different sampling rates (1500, 3000 and 4500 Hz) to provide an additional unseen sampling distribution class to the model’s performance. For each sampling rate, at least 50 signatures were collected to have a balanced dataset. The authors took this approach to create varied distributions of unseen datasets to train robust models. Furthermore, with the varied “unseen” sampling frequency (4500 Hz), the model’s robustness to unseen conditions was also examined. Testing different sampling rates allowed us to not only conduct device reliability testing but also collect data of different distributions from varying conditions. However, evaluating the model performance for different sampling rates is potential future work.

The specifications of the training, validation and testing datasets are presented in Tables 1–3, respectively, and illustrate the variations in the gathered dataset. The pumps’ speed and voltage are included to show the variety of data distributions used in training the model, as it might influence the fault development and may create a variety of features for the model to learn. For example, ref. [35] suggested that the speed can be used to understand the pump performance degradation. They also argued that the induction motor condition can be estimated by spectral analysis of the motor current and voltage waveforms. Therefore, it is important to share this information for potential further research areas when there is even more data to independently investigate the effects of these data on fault development. The faults were manually identified by the pump maintenance company engineers based on their knowledge of the pumps.

**Table 1.** Training pump specifications.

ID	Fault	Voltage (V)	Speed (rpm)
Tr_F1	Stator winding	415	1385
Tr_F2	Stator winding	415	1385
Tr_F3	Excess vibr., bearing dmg.	400	2815

**Table 1.** *Cont.*

ID	Fault	Voltage (V)	Speed (rpm)
Tr_NF1	Non-faulty	415	1450
Tr_NF2	Non-faulty	415	1450
Tr_NF3	Non-faulty	400	1450

**Table 2.** Validation pump specifications.

ID	Fault Type	Voltage (V)	Speed (rpm)
Va_F1	Impeller dmg., excess vibr.	400	1420
Va_NF1	Non-faulty	400	1450
Va_NF2	Non-faulty	400	1450
Va_NF3	Non-faulty	400	2815
Va_NF4	Non-faulty	415	1445
Va_NF5	Non-faulty	415	1445

**Table 3.** Testing pump specifications.

ID	Fault Type	Pump Type	Voltage (V)	Speed (rpm)
Te_F	Cavitation	Booster pump	425	2800
Te_NF	Non-faulty	Booster pump	415	2860

As noted in our prior work [17], there are very few publications that used real data, and among those publications, the dataset size is pretty small, which presents an issue for developing a robust machine learning model. For this reason, data augmentation techniques were used to increase our dataset sixfold. The total list of these augmentations for the training dataset was vertical mirroring, horizontal mirroring, 90-degree clockwise rotation, 180-degree clockwise rotation and 270-degree clockwise rotation. During the augmentation, the original image was used to obtain five augmented versions, which were saved in the dataset. The dataset sizes can be seen in Table 4. It should be noted that the training dataset did not have any data that were sampled at a 4500 Hz frequency. Due to the scarcity of data, some of the 3000 Hz data from two pump sites were artificially noised. The data were used as part of the training and validation. However, all testing was performed only with data from real pumps.

**Table 4.** Dataset sizes for each sub-dataset where training dataset was augmented.

Dataset Name	Faulty Data	Non-Faulty Data	Total
Training	6348	5688	12,036
Validation	370	538	908
Testing	107	107	214

The booster test pumps' signatures were never used for training and validation and were only used for the final result reporting and comparison.

## 5. Results and Discussion

### 5.1. Results Comparison

Table 5 shows the precision, recall and accuracy for testing and validation datasets. As a reminder, the data for the validation dataset was collected from real pumps that were different from those used for training. Therefore, the results present the model's

performance against the unseen validation and testing datasets. Table 6 shows that the ResNet-34 model performed the best at a higher sampling frequency (i.e., 4500 Hz) for both faulty and non-faulty pumps, despite not being trained on that frequency. However, it was also observed that non-faulty pumps had comparatively worse accuracy than the faulty pumps.

For a further demonstration of the model's performance, the following metrics were used: precision, recall, accuracy and F1-score, as can be seen in Table 5. The datasets' sampling frequency for these experiments can be found in Section 4.3. These metrics are defined as follows: Precision =  $TP/(TP + FP)$ , Recall =  $TP/(TP + FN)$ , Accuracy =  $(TP + TN)/(TP + TN + FP + FN)$  and F1-score =  $2TP/(2TP + FP + FN)$ . These metrics can demonstrate the model's robustness against overfitting. With the higher than 80% F1-score, we believe our model was capable enough to produce reliable decisions. Figure 3 shows the DQ patterns for the faulty and non-faulty pumps. It can be seen that the faulty pump signatures used in our experiments were characterised by irregular and distorted shapes compared with the non-faulty pumps. Table 7 shows the confusion matrices for 1500 Hz, 3000 Hz and 4500 Hz data.

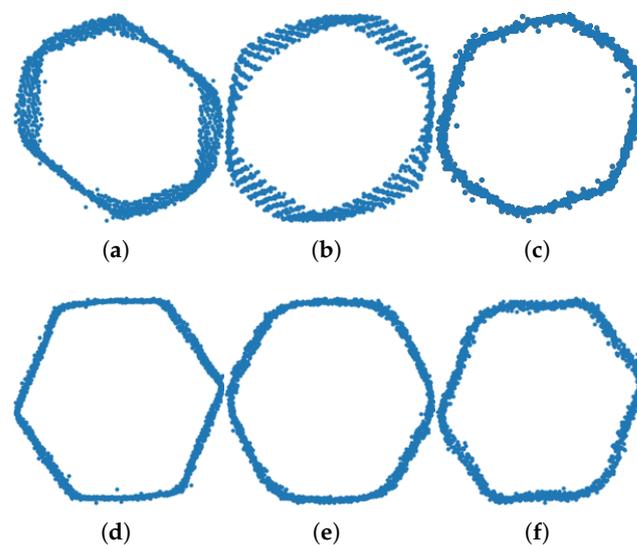


Figure 3. DQ patterns for faulty (a–c) and non-faulty pumps (d–f).

Table 5. Precision, recall and accuracy tables for validation and testing datasets.

Validation				
\	Precision (%)	Recall (%)	F1-Score	Accuracy (%)
Non-Faulty	87	95	91	88.33
Faulty	91	79	85	
Testing				
Non-Faulty	94	76	84	85.51
Faulty	80	95	87	

Table 6. Testing dataset size and results for each sampling rate and pump for ResNet-34.

Unseen Pump ID	Sampling Frequency (Hz) Correct Labels out of All Tests			Accu/Pump (%)
	1500	3000	4500	
Faulty Pump	35/37	17/20	50/50	95.33
Non-Faulty Pump	20/37	11 / 20	50/50	75.70
Accu/Samp. Rate	74.32	70.00	100.00	85.51

**Table 7.** Confusion matrices for 1500 Hz, 3000 Hz, 4500 Hz and the overall data.

1500 Hz	Faulty (Actual)	Non-Faulty (Actual)	3000 Hz	Faulty (Actual)	Non-Faulty (Actual)	4500 Hz	Faulty (Actual)	Non-Faulty (Actual)	Overall	Faulty (Actual)	Non-Faulty (Actual)
Faulty (Predicted)	35 (94.59%)	17 (45.95%)	Faulty (Predicted)	17 (85%)	9 (45%)	Faulty (Predicted)	50 (100%)	0 (0%)	Faulty (Predicted)	102 (95.33%)	26 (24.30%)
Non-Faulty (Predicted)	2 (5.41%)	20 (54.05%)	Non-Faulty (Predicted)	3 (15%)	11 (55%)	Non-Faulty (Predicted)	0 (0%)	50 (100%)	Non-Faulty (Predicted)	5 (4.67%)	81 (75.70%)

### 5.2. Discussion and Lessons Learned

The authors believe that the application of a convolutional neural network model to a non-conventional signal data type is a novel and relatively unique method. Despite its unorthodoxy, the model obtained decent results on both the validation dataset and the testing datasets. Furthermore, its relative performance on varying sampling rates also shows that it is a robust technique. The accuracy seems consistent with the reported fault detection accuracies in prior work [5,6,14], which ranged from 83.2% to 100%.

Although this study focuses on the machine learning aspect, the most significant challenges were faced in building, deploying and maintaining the infrastructure to collect the data. First, collecting faulty signatures from real pumps requires solving not only technical but also business dilemmas. Collecting faulty signatures requires a faulty pump or pump that is going to develop a fault soon, and thus, contradicts the customers' needs for continuous and non-faulty operation of their systems. The early replacements of these pumps is desirable to ensure smooth operation but can reduce the number of collected faulty signatures. Obtaining a sufficient number of faulty signatures to train the machine learning algorithm to correctly identify and label the faults has indeed been one of the major challenges. This is because modern pumps may operate for many years, and even after developing some faulty condition, may run under workable conditions for a long time. In addition, pumps can have multiple faults at the same time, which may complicate correct labelling and classification. To some extent, this could be mitigated by deploying data loggers on older pumps, which are more likely to fail, based on the maintenance engineer's knowledge.

A more reliable way to understand whether those motors/pumps are faulty or not is to intentionally damage the motor/pump to create the condition or investigate the pump/motor itself to look out for faults. While intentionally damaging the system is neither profitable nor exactly close to real life conditions, the close-up investigation of the system can render the pump/motor useless and create downtime for working systems. This is why we took extra care to limit such possibilities by using only pumps with known labels (faults), old unused pumps and intentionally damaging a small number of pumps. In the end, we obtained a very diverse dataset that contained a combination of real customer data and some intentionally damaged pumps from test benches.

At the hardware level, developing a system that reliably collected data 24/7 with minimal downtime is a challenging task. The hardware can be down, collect incorrect data or produce noisy readings. A system that works on a lab bench may become unreliable on site, and troubleshooting hardware and firmware issues can be time and resource consuming. From the design point of view, the hardware platform selection is determined by the cost, reliability and machine learning algorithm performance, which, in turn, requires a sufficient amount of data to be collected to design and tune correctly, resulting in a circular problem. This, to some extent, has been addressed in our project by an architectural decision of using lightweight IoT devices for collecting raw data and performing all processing, including machine learning and data storage remotely on the server. In the future, it may be interesting to experiment with distributed processing, where the fault detection is made locally by the device to minimise the reliance on the centralised server and connectivity problems.

Finally, at the machine learning level, the limited dataset size and the fact that data are unbalanced towards the non-faulty signatures may lead to bias and overfitting. Therefore, it is important to test the model under diverse scenarios, ensuring the model generalises

well for new data. The only reliable way to achieve this is testing the model with more data, which requires maintaining and operating the sensing infrastructure for an extended period of time, long enough for a sufficient number of pumps to fail to obtain the faulty signatures. Considering the relatively long lifespan of modern pumps, this may require several more years of data collection across several sites. This presents an immense practical and financial difficulty and is perhaps the main reason behind the scarcity of prior work involving real pump data. The contribution of this study is the development of such an infrastructure and offering an insight into the ML performance with real rather than simulated data.

Furthermore, given that the non-faulty pump/motor can become faulty over-time, the data that are used to train the model needs to be regularly checked to maintain the correctness. We also realized that the way the DQ images are resized, center cropped, etc., or the format they were saved/plotted vastly affected the model's performance. Therefore, it needs to be noted that the training on saved images vs. detecting on images on buffer needs to be configured as equally as possible.

## 6. Conclusions and Future Work

In conclusion, this paper presents a novel approach for centrifugal pump induction motor fault detection using data from real operational pumps deployed at various sites. This study reviewed the state-of-the-art residual network models on Concordia patterns and proposed an algorithm based on the image-based analysis of DQ patterns, which showed a promising 85.51% accuracy. The authors believe this is the first study that used a combination of transfer learning and DQ transformation to classify motor faults and a rare research paper that is based on real customer data. It is hoped that this study will provide valuable insights to other researchers on experimenting with image detection techniques on motor fault detection.

Potential future work includes expanding the training, validation and test datasets with data from more pumps, as well as predicting multiple fault possibilities instead of using a binary classification. Although the method was developed for centrifugal pumps, it would be interesting to explore the potential for other applications where alternate current (AC) at a known frequency is used to power electric motors. The DQ representation of the electric current signatures could be used to detect potential failures. Finally, the fault classification was performed in batch mode in this study. Processing the data in real-time and making classification online as samples arrive from the sensors may be beneficial in terms of latency and is potential future work.

**Author Contributions:** Conceptualization, C.E.S., V.V. and V.D.; methodology, C.E.S., V.V. and V.D.; software, C.E.S.; hardware: B.N., validation, C.E.S.; writing—original draft preparation, C.E.S.; writing—review and editing, V.D. and V.V.; supervision, V.V. and V.D.; funding acquisition, V.V., V.D., B.N. and J.N. All authors read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by Deckpro Pumps/Uptime Systems Ltd. and in part by the Innovate U.K. through a Knowledge Transfer Partnership (KTP) under grant 12028.

**Data Availability Statement:** The data collected and analysed in this study belongs to Deckpro Pumps/Uptime Systems Ltd. and can only be shared with their consent.

**Acknowledgments:** The authors are grateful to Usman Dar and Oliver Lewis for their help with the hardware development.

**Conflicts of Interest:** The fourth and fifth authors, namely, Barry Newton and Jake Newton, are employed by Deckpro Pumps/Uptime Systems Ltd. All five authors submitted a UK patent application GB2402869.8, containing the methodology and results of this study. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Capurso, T.; Bergamini, L.; Torresi, M. A new generation of centrifugal pumps for high conversion efficiency. *Energy Convers. Manag.* **2022**, *256*, 115341. [CrossRef]
2. Miljković, D. Brief Review of Motor Current Signature Analysis. *Croat. Soc.-Non-Destr. Test. J.* **2015**, *5*, 14–26.
3. Singhal, A.; Khandekar, M. Bearing fault detection in induction motor using motor current signature analysis. *Int. J. Adv. Res. Electr. Electron. Instrum. Eng.* **2013**, *2*, 3258–3264.
4. AlShorman, O.; Alkhatni, F.; Masadeh, M.; Irfan, M.; Glowacz, A.; Althobiani, F.; Kozik, J.; Glowacz, W. Sounds and acoustic emission-based early fault diagnosis of induction motor: A review study. *Adv. Mech. Eng.* **2021**, *13*, 1687814021996915. [CrossRef]
5. Fang, R.; Ma, H. Application of MCSA and SVM to Induction Machine Rotor Fault Diagnosis. In Proceedings of the 2006 6th World Congress on Intelligent Control and Automation, Dalian, China, 21–23 June 2006; Volume 2, pp. 5543–5547. [CrossRef]
6. Bacha, K.; Salem, S.B.; Chaari, A. An improved combination of Hilbert and Park transforms for fault detection and identification in three-phase induction motors. *Int. J. Electr. Power Energy Syst.* **2012**, *43*, 1006–1016. [CrossRef]
7. Valtierra-Rodriguez, M.; Rivera-Guillen, J.R.; Basurto-Hurtado, J.A.; De-Santiago-Perez, J.J.; Granados-Lieberman, D.; Amezcua-Sanchez, J.P. Convolutional Neural Network and Motor Current Signature Analysis during the Transient State for Detection of Broken Rotor Bars in Induction Motors. *Sensors* **2020**, *20*, 3721. [CrossRef]
8. Khan, T.; Alekhya, P.; Seshadrinath, J. Incipient Inter-turn Fault Diagnosis in Induction motors using CNN and LSTM based Methods. In Proceedings of the 2018 IEEE Industry Applications Society Annual Meeting (IAS), Portland, OR, USA, 23–27 September 2018; pp. 1–6. [CrossRef]
9. Dutta, N.; Umashankar, S.; Shankar, V.K.A.; Padmanaban, S.; Leonowicz, Z.; Wheeler, P. Centrifugal Pump Cavitation Detection Using Machine Learning Algorithm Technique. In Proceedings of the 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), Palermo, Italy, 12–15 June 2018; pp. 1–6. [CrossRef]
10. Tiwari, R.; Bordoloi, D.; Dewangan, A. Blockage and cavitation detection in centrifugal pumps from dynamic pressure signal using deep learning algorithm. *Measurement* **2021**, *173*, 108676. [CrossRef]
11. Irfan, M.; Glowacz, A. Design of a Novel Electric Diagnostic Technique for Fault Analysis of Centrifugal Pumps. *Appl. Sci.* **2019**, *9*, 5093. [CrossRef]
12. Perovic, S.; Unsworth, P.J.; Higham, E.H. Fuzzy logic system to detect pump faults from motor current spectra. In Proceedings of the Conference Record of the 2001 IEEE Industry Applications Conference, 36th IAS Annual Meeting (Cat. No. 01CH37248), Chicago, IL, USA, 30 September–4 October 2001; Volume 1, pp. 274–280. [CrossRef]
13. Toma, R.N.; Kim, J.M. Induction Motor Bearing Fault Diagnosis Using Statistical Time Domain Features and Hypertuning of Classifiers. In *Advances in Computer Science and Ubiquitous Computing*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 259–265.
14. Rapur, J.S.; Tiwari, R. Automation of multi-fault diagnosing of centrifugal pumps using multi-class support vector machine with vibration and motor current signals in frequency domain. *J. Braz. Soc. Mech. Sci. Eng.* **2018**, *40*, 1–21. [CrossRef]
15. Sivaprakasam, R.; Devaraj, D.; Rao, K. Fault detection in centrifugal pumping systems using neural networks. *Int. J. Model. Identif. Control* **2008**, *3*, 131–139. [CrossRef]
16. Dhomad, T.; Jaber, A. Bearing Fault Diagnosis Using Motor Current Signature Analysis and the Artificial Neural Network. *Int. J. Adv. Science Eng. Inf. Technol.* **2020**, *10*, 70–79. [CrossRef]
17. Sunal, C.E.; Dyo, V.; Velisavljevic, V. Review of Machine Learning based Fault Detection for Centrifugal Pump Induction Motors. *IEEE Access* **2022**, *10*, 71344–71355. [CrossRef]
18. Ameid, T.; Menacer, A.; Talhaoui, H.; Azzoug, Y. Discrete wavelet transform and energy eigen value for rotor bars fault detection in variable speed field-oriented control of induction motor drive. *ISA Trans.* **2018**, *79*, 217–231. [CrossRef] [PubMed]
19. Becker, V.; Schwamm, T.; Urschel, S.; Antonino-Daviu, J.A. Fault Investigation of Circulation Pumps to Detect Impeller Clogging. *Appl. Sci.* **2020**, *10*, 7550. [CrossRef]
20. Szabo, L.; Tóth, F.; Ernő, K.; Fekete, G. An Overview on Induction Machine's Diagnosis Methods. *J. Comput. Sci. Control. Syst.* **2008**, *1*. Available online: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=caafac443506c75cfbba3790429f924e7076c30d> (accessed on 8 April 2024)
21. Merizalde, Y.; Hernández-Callejo, L.; Duque-Perez, O.; López-Meraz, R.A. Fault Detection of Wind Turbine Induction Generators through Current Signals and Various Signal Processing Techniques. *Appl. Sci.* **2020**, *10*, 7389. [CrossRef]
22. Henao, H.; Capolino, G.A.; Fernandez-Cabanas, M.; Filippetti, F.; Bruzzese, C.; Strangas, E.; Pusca, R.; Estima, J.; Riera-Guasp, M.; Hedayati-Kia, S. Trends in Fault Diagnosis for Electrical Machines: A Review of Diagnostic Techniques. *IEEE Indus. Electron. Mag.* **2014**, *8*, 31–42. [CrossRef]
23. Şen, M.; Kul, B. IoT-based wireless induction motor monitoring. In Proceedings of the 2017 XXVI International Scientific Conference Electronics (ET), Sozopol, Bulgaria, 13–15 September 2017; pp. 1–5. [CrossRef]
24. Cano-Ortega, A.; Sánchez-Sutil, F. Monitoring of the Efficiency and Conditions of Induction Motor Operations by Smart Meter Prototype Based on a LoRa Wireless Network. *Electronics* **2019**, *8*, 1040. [CrossRef]
25. de Carvalho, D.P.; Silva, F.B.; Vanço, W.E.; da Silva Gonçalves, F.A.; Bissochi, C.A.; Monteiro, R.V.; Guimarães, G.C.; de Andrade, D.A. A method for real-time wireless monitoring of the efficiency and conditions of three-phase induction motor operation. *Electr. Power Syst. Res.* **2018**, *157*, 70–82. [CrossRef]

26. Sridhar, S.; Rao, K.U.; Nihaal, M.S.; Shetty Aashik, A.K. Real time wireless condition monitoring of induction motor. In Proceedings of the 2016 IEEE Industrial Electronics and Applications Conference (IEACon), Kota Kinabalu, Malaysia, 20–22 November 2016; pp. 173–178. [[CrossRef](#)]
27. Kunthong, J.; Sapaklom, T.; Konghirun, M.; Prapanavarat, C.; Ayudhya, P.N.N.; Mujjalinvimut, E.; Boonjeed, S. IoT-based traction motor drive condition monitoring in electric vehicles: Part 1. In Proceedings of the 2017 IEEE 12th International Conference on Power Electronics and Drive Systems (PEDS), Honolulu, HI, USA, 12–15 December 2017; pp. 1184–1188. [[CrossRef](#)]
28. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
29. Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870. [[CrossRef](#)]
30. Ren, W.; Jin, N.; OuYang, L. Phase Space Graph Convolutional Network for Chaotic Time Series Learning. *IEEE Trans. Ind. Inform.* **2024**, 1–9. [[CrossRef](#)]
31. Ren, W.; Jin, Z. Phase space visibility graph. *Chaos Solitons Fractals* **2023**, *176*, 114170. [[CrossRef](#)]
32. Chinthamu, N.; Karukuri, M. Data Science and Applications. *J. Data Sci. Intell. Syst.* **2023**, *1*, 83–91. [[CrossRef](#)]
33. Wang, Z.; Gao, J.; Kan, H.; Huang, Y.; Tang, F.; Li, W.; Yang, F. ResNet for Histopathologic Cancer Detection, the Deeper, the Better? *J. Data Sci. Intell. Syst.* **2023**. [[CrossRef](#)]
34. Uptime Systems Limited. 2023. Available online: <https://www.uptime.uk.com/> (accessed on 1 April 2024).
35. Harihara, P.P.; Parlos, A.G. Fault Diagnosis of Centrifugal Pumps Using Motor Electrical Signals. In *Centrifugal Pumps*; Papantonis, D., Ed.; IntechOpen: Rijeka, Italy, 2012; Chapter 2. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.