DIANOMIC

White Paper

Using vibration data monitoring to improve the reliability and efficiency of industrial centrifuges

"It is impossible for humans to understand the real time interactions between components of a complex machine... **The best way to do this is machine learning (ML)"**

Daniel Riddoch
May 2, 2023

1 Introduction and background

Condition monitoring of machinery used in manufacturing is essential to ensure the proper quality of product, reliability of supply, and safety of the workforce. Rotating equipment provides extra dangers as large forces are created on a comparatively physically small scale. There are numerous potential methods to track machine health. These can include PLC (programmable logic controller) monitoring, normal and thermal imaging, emission monitoring, and monitoring vibration data.

In specific cases, all of the above may be extremely good ways to monitor machines. However, monitoring of vibration data is the most common, richest, and broadest method[1]. Proper monitoring of vibration data can imitate the use of two human senses, sound and touch, as well as providing lots of information not available to human operatives, such as order multiple effects, resonance, and development of new patterns[2,3] . These can be used to identify machine problems such as: imbalance, looseness, misalignment, and bearing wear. In short, vibration is the gold standard for machine health monitoring, or as Hans-Dieter Bischof, Director of Research and Development at SKF GmbH, said: "Vibrations are a window into the health of machines, and through that window, we can see opportunities for improving machine performance and reducing maintenance costs."

Dianomic specialise in data pipelines and edge ML. What this means is that Dainomic can deliver the infrastructure needed to instrument machines, deal with the data created, process the data at the edge using machine learning and finally present that data in various forms to action the insights gained by the data processing. This flexible and scalable infrastructure is all controlled and managed using FogLAMP, and this provides the method for revolutionising the manufacturing sector, as this technology can be applied to all machines within a plant that can be instrumented, allowing for remote or even autonomous control of the plant. The vibration uses discussed here are an example of one specific way in which this technology can be utilised, although the versatility of this technology is another of its strengths, we focus here on the vibration use case.

In this paper, we will discuss the ways in which effective monitoring of vibration data can benefit industrial users. The simplest way to do this is to consider a case study, a model solution which Dianomic have created together with Archer Daniels Midland (ADM). ADM is in the business of producing vegetable oil and lecithin from raw produce, in this case, soy beans. As part of the process, the beans are crushed to remove their oil, the product is then filtered to remove the solids and the resulting mixture is spun in a centrifuge to separate the light oil, the lecithin oil and remaining solids. These centrifuges are critical to the continuity, quality and safety of production. The efficacy and efficiency of these machines can be affected by a number of factors, including but not limited to, the build up of solids, wear on the moving parts, and timely maintenance without removing the production capacity. ADM wanted to improve the performance of their manufacturing facility, and engaged Dianomic Systems Inc. to monitor their centrifuges vibrations using relevant sensors and data acquisition within the FogLAMP system. FogLAMP is an Industrial Data pipeline and Edge application platform. FogLAMP uses a pluggable microservices architecture to connect all asset data, analyze the data and securely integrate to the existing systems that inform operators, engineers, maintenance teams, and data scientists, see section 4 for more details.

Assembled machines such as centrifuges contain hundreds, if not thousands, of interactions between components. These can take various forms, and need to be studied in different ways. Significant literature could be devoted to the static contacts[4,5], dynamic effects[6,7], or modelling the state of stress[8,9]. Modelling these across different length scales, interacting with each other, by traditional means[10] is practically impossible and would be a life's work for a single machine. Numerical methods offer some shortcuts, but sacrifice adaptability and accuracy for speed, which is still not enough to keep up with the rapidly changing conditions of reality[11].

Realistically, it won't be possible for humans to understand the interaction between components of a machine in real time, and without this, we can't tell what will happen next. Furthermore, the only solution is trying to spot patterns and detect changes that might be a symptom of a problem. The best way to do this is machine learning (ML), which allows for a more versatile, bespoke, and sensitive approach. Fundamental to the ML life-cycle is collecting and labeling data for the data scientist to build and test models.

To collect accurate data, there are two potential sources. The first is to use accelerometers. The details of how to choose and place accelerometers are discussed in section 3.1. Accelerometers give direct vibration information, and can measure and detect unusual reactions. The second way to collect data is to take it directly from the machine control system. This will be more accurate and cheaper than accelerometer data, and is crucial to contextualise all data. The control system data will also include the draws on the machine, which can be calibrated to reflect the loads and therefore the status of the machine. Combining accelerometer and control system data sources, we are able to get a full accurate picture of the machine state in previously unrealised detail, speed, and efficiency.

2 Detailed case study

As already mentioned, a case study will be presented, to demonstrate plainly the advantages of monitoring vibration data. This will serve as a reference when discussing details of the machines, methods, and acquisition procedures. In this case, we focus on a specific centrifuge, in a specific part of the production process. However, the methodology is applicable to all kinds of centrifuges, rotating machines or similar procedures. Obviously, each case will be slightly different and care should be taken in applying the techniques described here, but the principles are sound.

2.1 Detailed case study

The centrifuge is part of a production line producing vegetable oil from soybeans. The beans are delivered to the plant, then crushed and oil is separated from the bean. This then gets filtered, hydrated and then spun to separate oil from the gums (lecithin oil) known as degumming. Further refinement of the oils takes place in further centrifuges, but the use case we are concerned with is a degumming centrifuge.

This means that the input to our centrifuge is a mixture of all the liquids (we consider this to be two different types of oil) and a few residual solids which get through the filtration process. The purpose of this centrifuge is two fold; we wish to separate the two different types of oil, these are of different densities and can be separated by centrifugation; and we wish to remove any solids which have passed through the filter. All three of these phases; lecithin,light oil and meal(dried solids, sold as animal feed); are valuable to different degrees, although all lose value if the product is impure.

2.2 The centrifuge

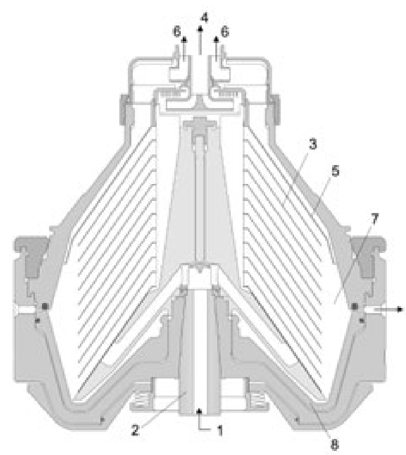
The centrifuge that has been used for this project is an Alfa-Laval PX-100. This is a popular industrial sedimentation centrifuge of the disc stack type. To improve the efficiency of the separation process, the Alfa-Laval PX-100, and other centrifuges of this type, contain a series of stacked conical plates, which add surface area for the deposition of solids, which eventually will be flung from the plates to the edge of the bowl. The increased surface area for sedimentary deposition meaning that the process is more efficient, faster, and produces higher quality product.

Operating principles

The oil to be separated is fed (1) into the separator bowl from the bottom through a hollow spindle (2) and enters the disc stack (3).

The heavy phase and heavy sludge are forced towards the periphery of the bowl, while the light phase flows towards the centre of the bowl, from where it is pumped out (4) for further processing. The heavy phase is led over a top disc (5) into a chamber where an adjustable paring device pumps it out of the separator (6).

Sludge collects in the sludge space (7) and is discharged intermittently and automatically. The discharge is achieved by a hydraulic system which at preset suitable intervals forces the sliding bowl bottom (8) to drop down, thus opening the sludge ports at the bowl periphery. The sludge is collected in the frame and leaves the centrifuge via a cyclone.



Typical bowl drawing for a solids ejecting hermetic centrifuge. Drawing details do not necessarily correspond to the centrifuge described.

Figure 1: Diagram of a typical centrifuge and flow explanation[13]. Credit - Alfa-Laval.

Figure 1 shows a diagram of the centrifuge, which shows its arrangement of discs as well as showing the flow of the product through the machine. The filtered oil is passed into the centrifuge through a hollow spindle at the bottom (arrow 1), and passes into the main chamber. The solids are heaviest and are flung to the edge of the bowl (arrow 7), the two phases of oil are left centrally. The lighter oil flows up the conical discs because of the pressure built up within the bowl, this is ultimately pumped out of the hollow spindle atop the machine (arrow 4). The heavier oil, known as lecithin oil does not flow up the discs, but is pumped out of the top of the machine through an annulus (arrow 6). The solids which accumulate are shot out using centripetal force when the two halves of the bowl separate, in a procedure known as a shoot, more information is presented on this in section 2.3, this happens at the unnumbered arrow on the right of the diagram.

The lecithin oil is most valuable and can be sold at prices around $1350 per metric ton, the light oil is sold for $1200 per tonne, and the solids can be dried and sold as meal for $550 per tonne. When the shoot occurs, it is inevitable that oil will be expelled along with the solids, as the solids are suspended in liquid. This oil cannot be recovered, so one of the aims therefore is to minimise the number of shoots and optimise their duration to minimise oil loss and solid accumulation within the bowl, which reduces the centrifuge efficiency

2.2.1 Drive mechanism

The centrifuge works using an in-built electric motor, whose rotational axis is set perpendicular to the rotational axis of the centrifuge. A worm gear then transfers this rotation to the correct axis, then driving the spindle through which the product is pumped in, this is connected to the stacked discs.

The Alfa Laval PX100 centrifuge is equipped with a planetary gear system that provides variable speed control[12]. The planetary gear system allows the bowl speed to be adjusted to achieve the desired level of separation performance. The planetary gear system consists of a sun gear, several planet gears, and a ring gear. The sun gear is driven by a motor and transmits power to the planet gears, which are mounted on a carrier that rotates around the sun gear. The planet gears mesh with the ring gear, which is attached to the centrifuge bowl, causing it to rotate.

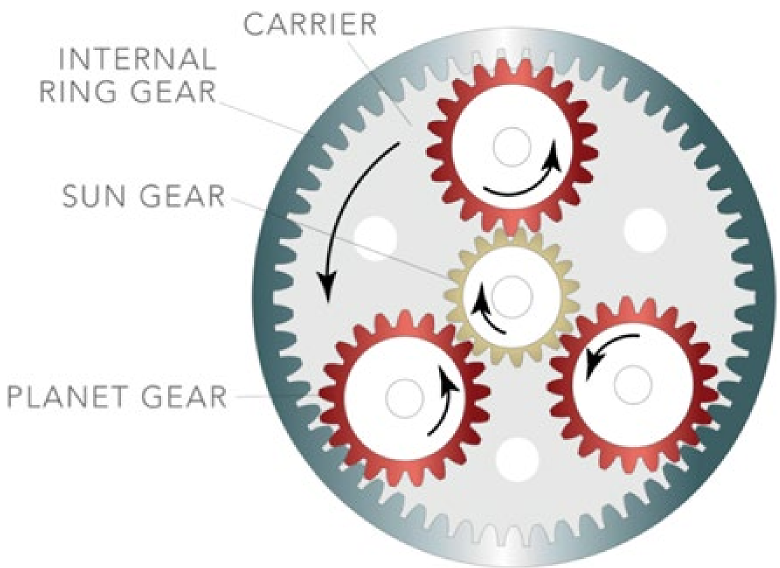


Figure 2: Diagram of a planetary gearbox.[14]

The planetary gear system provides a reduction of bowl speed and can be adjusted to optimize the separation performance for different process conditions. The bowl speed is typically controlled by a variable frequency drive (VFD), which adjusts the motor speed to achieve the desired bowl speed. A hollow spindle is used to spin the stacked discs, bowls and associated valves as is necessary.

2.3 Maintenance procedures

Having covered the purpose of the centrifuge, and how it works, we must now understand the maintenance procedures currently in place, the maintenance that is possible or viable to undertake and the effect it has. There are three maintenance procedures which are regularly carried out by ADM; the first of these is known as a shoot, during which the two halves of the bowl separate and the accumulated solids, along with some of the partially refined oil; the second is known as Lean in Place(CIP), where the production is stopped and water is fed in to the centrifuge to rinse and clean the bowl.; the third is a full service of the machine, which involves halting production, dismantling the machine and fully cleaning the internals of the machine. One of the objectives being discussed here is to minimise the number of maintenance procedures required. Furthermore, by optimising shoot intervals, yields increase and maintenance expense decreases.

During the shoot procedure, the bowl is split into two halves by releasing pressure in the chamber below the lower half of the bowl. The momentum of the solids will then carry these out of the bowl. Similarly, some oil will be ejected along with the solids. As the split in the bowl is fairly low in the bowl, most of the oil ejected will be lecithin oil from the heavy phase, but as this is before the refinement of the product, it is reasonable to expect a mixture of the light and heavy phases. The solids and oil are collected, dried and sold as meal, and as already mentioned oil is lost during this procedure. During a CIP event, water is used to clear the inside of the centrifuge bowl of all remaining solids. This is generally more effective than a shoot procedure, as any solids stuck anywhere other than around the middle of the bowl can be removed, and a greater proportion of the solids are removed in any case. However, CIP is more disruptive than a shoot event, because the flow of oil must be stopped, and oil left in the centrifuge will be lost, and residual water will likely affect the purity of the oil for a period of time after the CIP cleaning.

A service of the machine is extremely disruptive to the production, as the continuous flow of oil is interrupted. The process also requires skilled operatives to carry out this maintenance, and finally, the stopping, and restarting of the machine places additional wear on components. Another of the objectives of the monitoring therefore is to delay the services as long as possible, while still ensuring the safety of the machine, and the quality and efficiency of the separation process.

3 Data collection

Having discussed the case study, and understood some of the challenges presented and the motivations for improving the machine's efficiency, let us move in to consider data acquisition. We have already mentioned that there are lots of potential ways to collect data, however, many of these are not applicable or viable for an industrial vibration monitoring set up. This could be because vibration is not itself being monitored, or because the measurement equipment involved is extremely expensive, or too delicate for industrial application and better suited to laboratory work.

As such, we can use two different complementary methods of data acquisition. First, we discuss using piezoelectric accelerometers, then we consider using data from the Programmable Logic Controller (PLC) which controls the machine itself, but also measures key parameters.

3.1 Accelerometer data

The rotation of the centrifuge bowl causes vibrations which are detectable at the centrifuge casing. The vibration is suitable to characterise the state of the centrifuge as it is affected not only by the composition of the oil being processed in the centrifuge, but also by other technical problems like a misalignment of the centrifuge bowls.

Piezoelectric accelerometers work by utilising a small piezoelectric crystal and a signal amplifier to convert mechanical work into electrical signals. This works by sandwiching a piezoelectric crystal between two metal plates. When a force is applied to compress the crystal, the charges inside the crystal move, the fine balance of electrical charges is disturbed, and a net charge results, as illustrated in figure 3. The charges produced are small, but can be measured by sensitive electronic equipment. This can then track the changes in voltage across the crystal, which can be correlated to the force, and consequently the acceleration, experienced by the crystal. This is given by

$$\sigma_{ij} = d_{ijkl} E_{kl} + e_{ijkl} T_{kl} ,$$

where σ_{ij} is the stress tensor, d_{ijkl} and e_{ijkl} are the piezoelectric constants, E_{kl} is the electric field, and E_{kl} is the strain tensor.

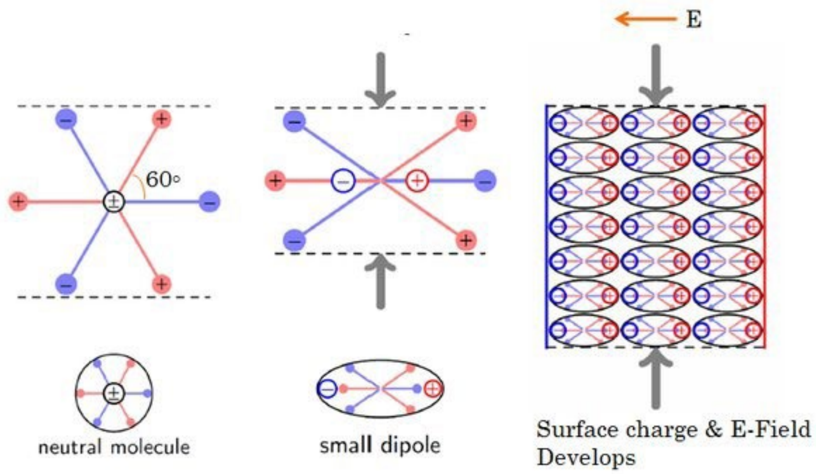


Figure 3: Illustration of dipole charge movement in a piezoelectric crystal.
This effect can be measured to produce accelerometers.[15]

3.1.1 Accelerometer positioning

Data is collected by four accelerometers mounted to the centrifuge casing as well as the motor as shown in Figure 4. The positions to mount the accelerometers to the centrifuge casing were chosen to detect not only translations in all three dimensions, but also a possible deformation of the bowl or casing.

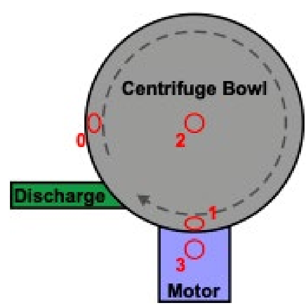


Figure 4: Positioning of accelerometers in top-down view of the centrifuge.

The translation of the bowl could be detected by a single triaxial accelerometer as shown in Figure 5(a). However, a deformation of the centrifuge bowl as shown in Figure 5(b) might falsely be interpreted as a translation. The chosen positions of the accelerometers provide enough information to detect deformations as shown in Figure 5(b).

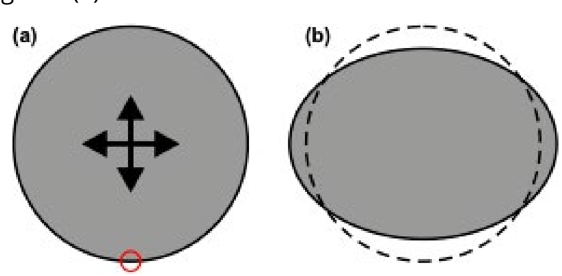


Figure 5: (a) Centrifuge translation detectable with a single triaxial accelerometer and (b) deformation of centrifuge bowl not detectable with a single accelerometer.

3.1.2 Accelerometer type

Two different accelerometer types are used on the same prototype centrifuge, the PCB623C00 and the PCB623C01. Both are manufactured by PCB Piezotronics and are uni-axial high frequency accelerometers with ceramic sensing elements. The key product specifications for both variants are compared in Table 1.

	623C00	623C01
Sensitivity (±5%) in mV / ms^{-2}	1.0	10.2
Frequency range (±5%) in Hz	2.4 — 8000	2.4 — 8000
Frequency range (±10%) in Hz	1.7 — 10000	1.7 — 10000
Frequency range (±3dB) in Hz	0.8 — 15000	0.8 — 15000
Measurement range in g	±500	±50

Table 1: Table comparing key values for the two existing accelerometers.

The accelerometers collect acceleration data in g. They differ only in the measurement range and the sensitivity. Measurements on the centrifuge have shown that the smaller measurement range of the PCB623C01 accelerometer is sufficient for the data collection, and hence its higher sensitivity became the deciding factor for the accelerometer choice. Other accelerometers with similar performance are available, the choice is down to the individual user.

3.1.3 Sampling rate

The accelerometer choice places constraints on the ability to choose the sampling rate. A higher sampling rate will present a number of advantages; the data is likely to be more accurate with less chance of peak cut off or artificially creating beating effects; the frequency resolution is better, allowing a fuller understanding of all frequency components, rather than only observing lower order effects of high frequency vibrations; and, a fuller understanding of transient, or short term, behaviours, as can be seen in figure 6.

However, high frequency sampling is limited by equipment performance, and will produce a large amount of data. These factors can combine to increase the cost of the equipment necessary to measure and store these data, these costs can rapidly escalate as they can be encountered every time the data is collected, transferred and stored. Higher frequency sampling can also lead to a greater level of noise, especially if the capabilities of equipment are being stretched.

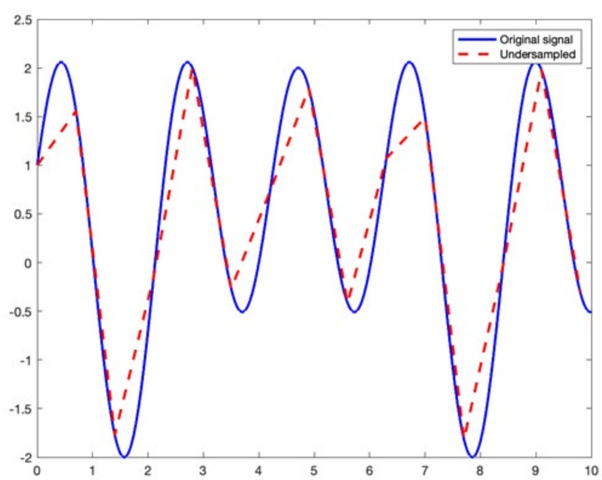


Figure 6: Example of under sampling, showing peak cut-off, a small amount of beating, and missing certain frequency components.

Any choice of sampling rate is therefore necessarily a compromise. It is necessary to choose a sufficiently large value that all the detail of the signal can be captured for proper analysis, whilst working within the constraints of the equipment, and minimising the sampling rate to reduce costs. To this end, a samplingvrate of 8000 Hz was chosen for the centrifuge. This allows the greatest accuracy of measurement whilst still capturing as much detail as possible.

3.2 Control system data

The centrifuge control system, commonly referred to as the PLC, generates and monitors in near real time several different streams of data, which are of use individually, but can, when combined with accelerometer data, give a clearer picture of the state and health of the centrifuge.

3.2.1 Labelling data

Perhaps the simplest way that control system data is used is to label the data sets collected, collocate the PLC and accelerometer data and easily identify different data sets. There are several obvious advantages to this, firstly it allows the data collected to be intuitively catalogued and labelled. This allows the data to be easily accessed by operatives to further investigate problems, check machine histories, and verify the proper operation of the ML procedures.

The data will also enable us to determine what the machine is doing during data collection. Most simply, we can determine if the machine is stopped, starting up, running normally or stopping, as can be seen in figure 7. But, we can also determine the length of time since the last shoot, or service, how long the machine has been running continuously, and link the machines performance to data collected in other areas through the plant, for example at the hydration tank.

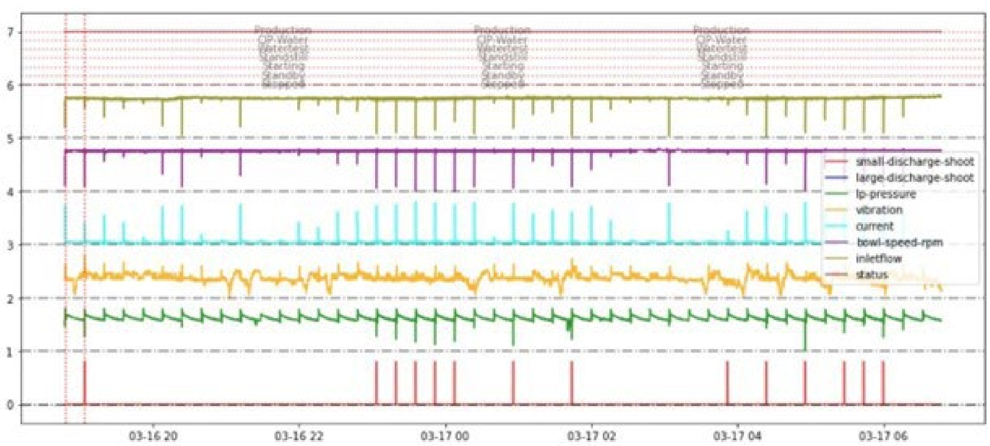


Figure 7: Figure showing different data collected from the PLC, including labelling (top), electrical loads (centre, blue) and the accelerometer data (centre, yellow).

Understanding the state of the machine will allow for comparisons with other data sets with the machine in similar condition, which allows us to monitor a changing normal, and account for expected changes which do not present a problem in themselves.

3.2.2 Machine parameters

The PLC labelling can extend further to give us information about the machine state. The simplest and yet one of the most important pieces of information is information about the speed of the machine. The machine speed would be expected to show up in the measured vibration data, and we expect to see several machine speed order effects, which will appear at whole number multiples of the machine operating speed, as shown in figure 8.

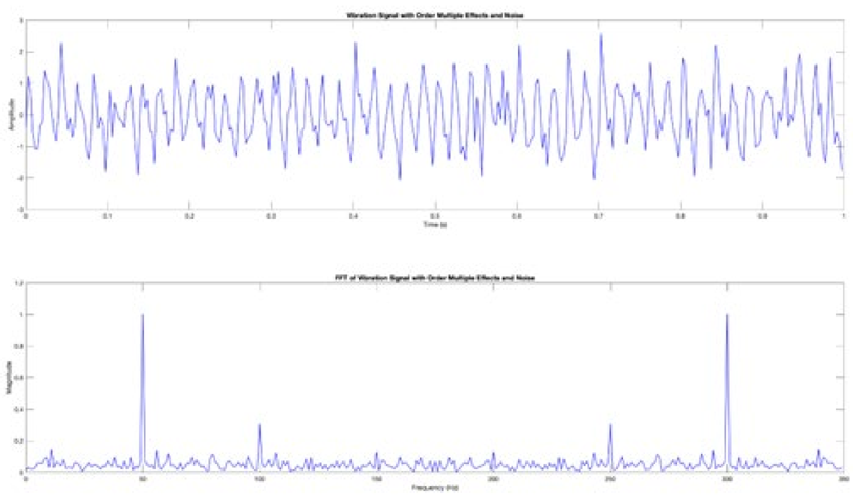


Figure 8: Plot showing how a vibration signal may contain effects at orders of a single value.

The PLC will also collect and store data about the input and output flows to the machine. Whilst the masses of the through flows are not measured, the duration of time the valves are open will be indicative of the amount of material in the bowl. This changes the mass of the centrifuge and consequently will change its natural vibration frequencies.

3.2.3 Electrical loads

The last bit of PLC data that we will utilise is the electrical loads. The PLC monitors the voltage and current draw of the motor. Increased draws on the motor can be seen in increased electrical draws. There are a few potential reasons for increased draws, which may include, build-up of solids, wear within the drive system or centrifuge bearings, and excessive quantity of material within the centrifuge bowl.

It is not possible to distinguish between these factors based purely on the electrical signal, which is why it is necessary to combine the PLC data with accelerometer data, as illustrated by figure 7. Similarly, accelerometer data alone cannot isolate an unusual vibration as being based within say the gearbox, or within the centrifuge itself. As such, both data collection methods are greater than the sum of their parts when intelligently combined.

4 Data processing

Having collected and collated the data, we must now make good use of it. It is worth considering at this point, the sheer volume of data created by the acquisition methods just described; choice of 8KHz sampling rate brings into focus the large amount of vibration data that needs to be processed. To get a sense of the data volume in this case:

Data	Data volume (per accelerometer)	Data volume (4 accelerometers)
1 sec	8000 x 64-bit floats = 64 KB	64 x 4 = 256 KB
1 min	3.75 MB	15 MB
1 hour	225 MB	900 MB
24 hours	5400 MB (5.27 GB)	21600 MB (21.09 GB)

Table 2: Table showing the volume of data created by the accelerometers mounted on a single centrifuge, as described in section 3.1.

Evidently, we cannot send all this data for every possible machine to a single location, and dealing with such quantities of data would be slow and unwieldy. A simple solution is to only collect a subset of the data, for example 2 hours worth of data in a 24 hour period, and with compression techniques, this can be compressed to a file size of just 360 MB. However, the need to transfer this quantity of data during operational phases can be obviated using the edge processing and ML capabilities of FogLAMP.

4.1 FogLAMP and edge processing

One of the first choices to be made is where the processing should occur. Lots of computer power has traditionally been required for these types of operations, and this means that the data must be sent to a remote location for processing. This introduces problems with secure, reliable, fast, and cheap transfer of large quantities of data. In recent years, the computing power of in situ hardware has increased significantly, and there have been simultaneous improvements in the efficiency of such programs.

These factors combine to mean that, we are now able to consider the possibility of applying pre-developed ML models at the industrial edge. This removes the need for permanent data transfer equipment and will allow a faster evaluation of the machine health to be made. The transfer of some data will still happen, as the model cannot yet be trained at the edge, however, this opens up the ability to integrate the data processing into the machine control system, meaning collection of PLC data will be much easier, and a single system could be used to monitor and control the machine.

The transfer, control, management and processing of this data is where Dianomic's expertise lies, and the FogLAMP system has all the necessary tools to facilitate this. FogLAMP is designed to collate data from different sources, in the nomenclature this is data acquisition from its 'south.' Distinct processes can then be carried out, known as filters, which can produce data for other filters to use, or for data distribution. FogLAMP can then send data across many different protocols, to many destinations, this is known as the FogLAMP's 'north.' This is explained in figure 11, this covers two different things; the first of these is the ML operations pipeline, which covers the top right branch, where training data is collected and the model trained; the second is the operations pipeline, where the model is applied, this can also be seen in figure 19.

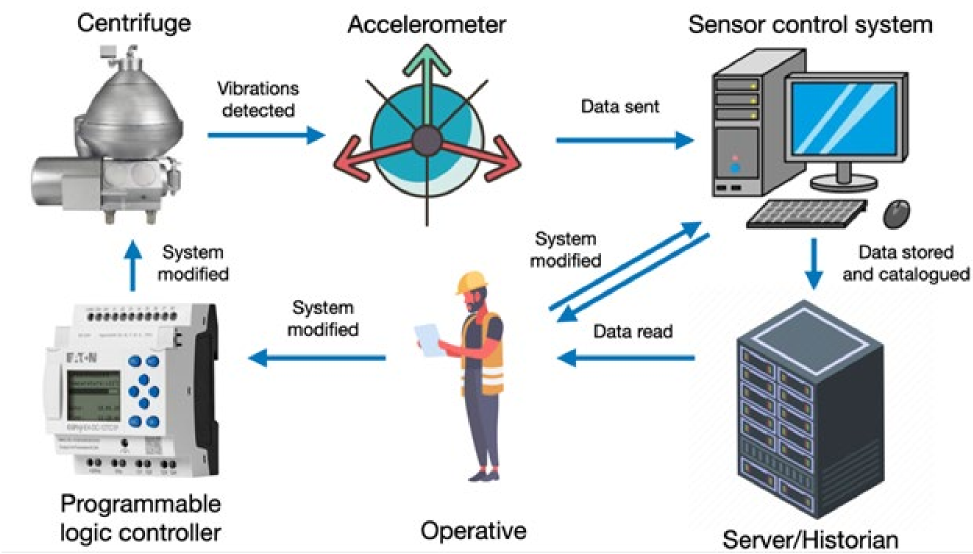


Figure 9: Cartoon showing traditional data monitoring flow.

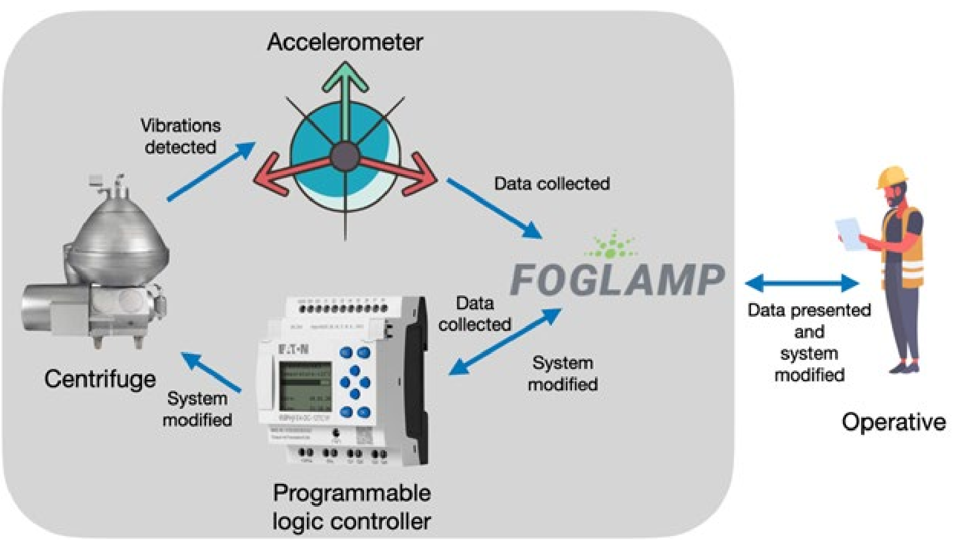


Figure 10: Cartoon showing data monitoring flow using FogLAMP.

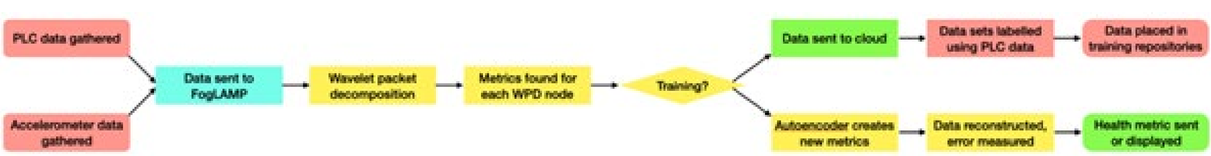


Figure 11: Flow chart showing the movement of data. Blue represents FogLAMP south, green FogLAMP north, yellow is within FogLAMP and red is external to FogLAMP.

4.2 Data handling

Having collected data, we must now start to use it efficiently. The first thing to consider is perhaps data security. At various points in the operating cycle, different types of data will have to be sent to and from the edge processing, which we may think of as at the machine itself. This may be different to previous systems, and this data is commercially sensitive, so appropriate security protocols should be in place.

The data pipelines can manage their own internal security, however when it comes to the terminations of the pipeline FogLAMP no longer has total control over the security mechanisms available and must work within the constraints of the external endpoints. Either the device the data is being collected from or the system to which the data is being sent. These will almost certainly not offer the same set of options as each other for authentication or encryption of the data, especially given the many to many nature of the data flows through a single FogLAMP. Data must be decrypted from whatever, if any encryption mechanism the data provider offers and then re-encrypted on egress from FogLAMP. That egress may be to multiple destinations each with their own set of encryption mechanisms available.

Typically, FogLAMP will authenticate with the destinations using the destinations preferred authentication mechanism. The same data, possibly from multiple sources, may be sent to multiple destinations using the authentication appropriate for that destination.

Now having the data to hand, we begin to consider how to use it. Perhaps the first thing to consider is the signal to noise ratio, and what efforts can be made to improve this and provide clearer signal data. There are several filters which may be applied, such as a moving average filter, which discard values beyond the standard deviation from a rolling average, this can work well to remove extreme points, but does not work well with high frequency data compared to the sampling rate, similarly auto-regression can be used but may struggle with data having frequency components of the same order of magnitude as the Nyquist frequency. Low, high and band pass filters, Kalman filters, or Wiener filters could also be used.

Raw vibration data is notoriously hard to understand, and it is normal to wish to investigate particular frequency bands or use these to identify different features. For this reason, it will be useful to consider the frequency space. There are several different ways to do this, which depend on the pattern we expect of the data.

If, across the sampled time, we do not expect the data to change in character, and we are confident about this, the best representation of the various frequencies in the signal, is found using a Fourier transform. This is a very powerful tool, based on deconstructing the signal into an infinite sum of sine (imaginary parts) and cosine(real parts) waves. The Fourier transforms turns a plot of a signal in the time domain into one in the frequency domain; if the amplitude (given on the y-axis) is multiplied by a sine wave having the frequency given on the x-axis, and all these points are summed, the original signal will be reproduced. The amplitude of a certain frequency can be found by

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt,$$

where $f(t)$ is the original function, ω is the chosen frequency and $i = \sqrt{-1}$. This allows us to precisely identify the frequencies having impact on the signal, compare their effect and identify if two similar frequencies are having large effects, which may indicate a possible resonance.

However, Fourier transforms cannot account for any changes in the time interval. It is important for us to consider this possibility, as we may see a shoot occur, or some other instantaneous effect, and it is important that we can determine the machine health after these effects. For this reason, we wish to be able to understand the changes throughout the time sample. This can be achieved still using a Fourier transform, by using a Short Time Fourier Transform (STFT) which allows the time axis to be sliced and lots of individual Fourier transforms to be performed. This works by using a Hanning window, which is a type of step function which can be written as

$$h(t) = \begin{cases} 0.5 - 0.5\cos\left(\frac{2\pi t}{T}\right), & |t| \le \frac{T}{2} \\ 0, & \text{otherwise.} \end{cases}$$

This can then be translated through time by use of the translation parameter τ , giving the equation

$$F(\omega\tau) = \int_{-\infty}^{\infty} f(t)h(t - \tau)e^{-i\omega t}dt$$

This, however, is a crude approximation which cannot properly explain changes within its bands, which must be wide enough to attain a reasonable accuracy in the transform.

For these reasons, a wavelet decomposition is used. The advantage of this is that it allows a changing time and frequency resolution, which captures more detail, is more adaptable to changes in the signal, and removes any requirement to predetermine or anticipate results. It works by reconsidering the underpinnings of Fourier analysis. Fourier analysis works by decomposing any function into an infinite sum of sine and cosine waves. If we consider function space as a vector space, then Fourier analysis relies on the fact that sine and cosine are a basis of the space, under the operations of multiplication within the argument, and multiplication and addition outside the argument. This means that all functions can be created by adding various combinations of sine and cosine waves with arguments that are multiples of each other.

However, sine and cosine are not unique as a basis for the function space. Furthermore, because of their continuous nature, they are suited best to periodic functions. More accurately, sine and cosine are a perfect basis of periodic function space, and only an approximate basis of non-periodic function space. For this reason, it may be desired, if the signal you wish to decompose might not be strictly periodic, but instead evolve with time, to consider an alternative basis. This is where wavelets can be considered. Wavelets are essentially a short burst of different functions, which are zero valued elsewhere, as shown by figure 12. These wavelets can be used as a basis of all function space in the same way as sine and cosine.

The contribution of different wavelets can be found using the integral:

$$F(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} f(t)\psi * (\frac{t-\tau}{s})dt,$$

where ψ* is the complex conjugate of the wavelet function, and s is the wavelet scale parameter 1/f. From this, it is theoretically possible to draw a graph, similar to that drawn from the Fourier transform showing the contributions of different wavelets at different scale parameters. However, it is likely, especially in real data signals, that this would result in information loss due to not accurately capturing small amplitudes, these being drowned out by the larger amplitudes. It is therefore more common to simply show a reconstruction of the signal using wavelets, than to list these wavelet coefficients.

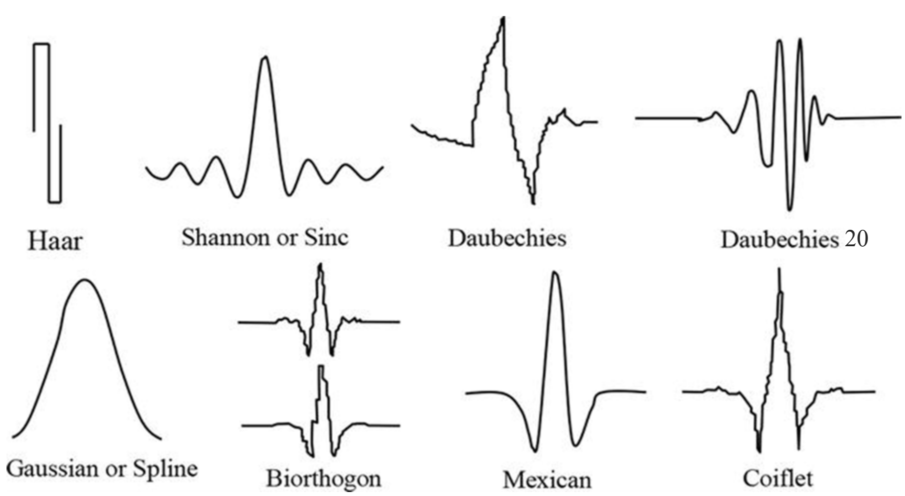


Figure 12: Example showing different types of wavelets[16].

Generating all these wavelet coefficients produces huge volumes of data[17], so in practice a discrete wavelet transform will often be used. This requires a choice of discrete values of s and τ for analysis to occur at. The choice of these values is not intuitive or easily understandable, so another technique called a wavelet packet decomposition is used. This breaks the frequency spectrum into chunks, or 'packets,' and based on the frequencies which bound the packet, an appropriate choice for s emerges. Similarly, the time axis can be 'packetised,' and an appropriate choice for τ emerges automatically. The coefficients for these packets can then be easily generated and used to decompose the signal. Some thought should be given to the "mother wavelet", the original form from which other wavelets are derived, and the decomposition level, which governs how many different scale factors are used simultaneously.

For a noisy sinusoid signal, like those encountered in vibration analysis, sinusoids, like the Daubechies 20 wavelet shown in figure 12, make a good choice as mother wavelet. Decomposition level is more difficult to generalise, and a greater decomposition level will give greater fidelity, but experience has shown that a decomposition at level 4 or 5 is almost always sufficient. Having these parameters, the regenerated signal can be compared to the original, as shown in figure 13.

In some other applications, some smoothing or thresholding of these coefficients is done, in an attempt to filter out noise. This was considered, but in our experience, especially when considering some smaller magnitude effects, the signal to noise ratio (SNR) is quite large. This means that any attempts to artificially remove the noise will likely lead to the creation of 'ghost effect' or numerical artefacts, created by mathematical processes not genuine effects.

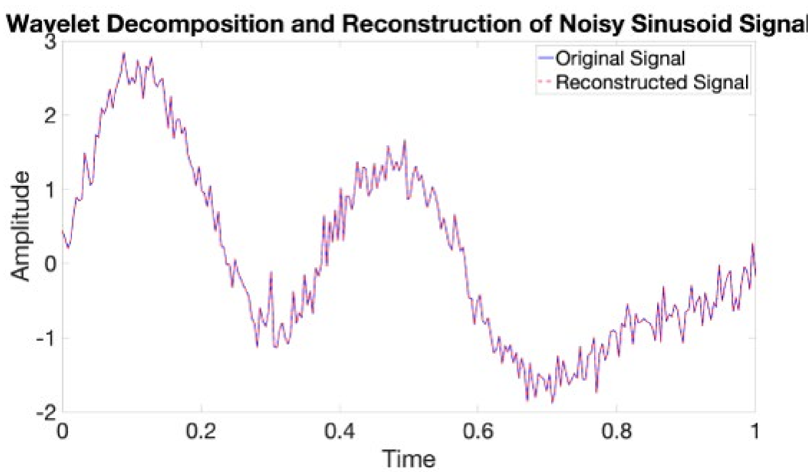


Figure 13: Example of a signal which can be decomposed and reconstructed using wavelet packet decomposition.

4.3 Data metrics

The statistics of these nodes is the data that we will use from now on. But the interpretation of them is important. Feeding raw data into ML algorithms will often fail to identify key features, trying instead to auto-correlate, which means overlaying the signals to compare them. It is perfectly possible for two healthy signals to look different, either because of order effects, noise, or a natural deterioration due to wear or solid accumulation which is not at a dangerous level.

To this end, a set of metrics are used to identify key features of the data. Some of these are fairly obvious, such as the mean(the mean of vibration data should be approximately zero, anything else would indicate a procession, which could be very dangerous — see figure 14) and standard deviation. Others are more obscure, such as Kurtosis, which measures peakedness of the signal's probability distribution around its mean — see figure 15.

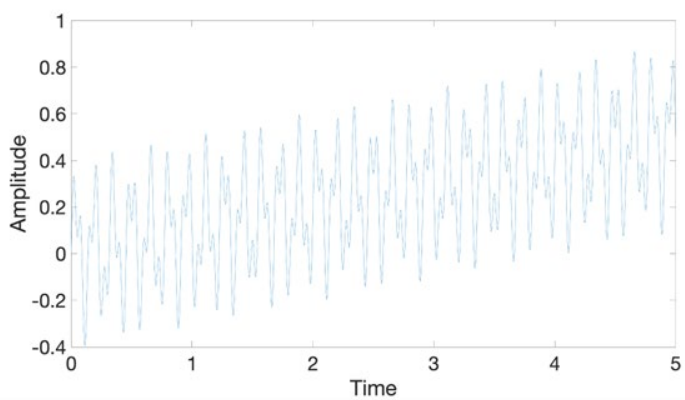


Figure 14: Sketch showing a processing sine wave.

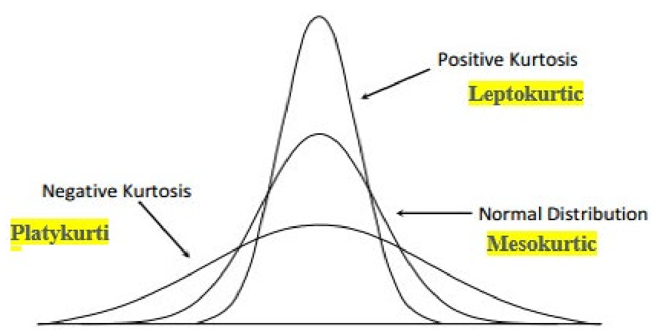


Figure 15: Sketch showing different values of kurtosis in a distribution.[18]

The details of the exact metrics, their uses, and how they are calculated is not particularly important, again, as long as these things are kept consistent. What matters is capturing the characteristics of the signal. Is the signal particularly 'jagged' in appearance, with large changes between data points, is the amplitude of vibration similar for each peak, and is the signal symmetric about its mean. There are various ways to calculate metrics to discern these characteristics.

It is not necessary to go into specifics because the point of using ML is the ability of the system to learn and create new metrics which are tailored to the data sets provided, see sections 5.1.2 and 5.3. This is the advantage of ML, because the system resulting from a well built ML model will be more specific, bespoke and tuned to detect defects in the machine in a way which is far more relevant than any generic system we may design.

5 Machine learning

As has been previously stated in the introduction to this document, it will not be possible to build an accurate mathematical model of such a large and complex machine as the Alfa-Laval PX100, or any similar machine for that matter. We have also discussed that numerical approximations to this do not fully account for this shortfall, sacrificing accuracy for speed, but in all probability, still not fast enough to be useful in condition monitoring. The only way in which we can usefully apply the collected data in order to meet the stated aims, is with a machine learning model.

5.1 Machine learning models

With an understanding of the data, let us briefly cover the concept of machine learning. What we mean by these is the process of using large sets of data to ‘train’ a mathematical model. What this means is that a basic model is built by the user, and the details, what we might think of as the coefficients of a line of best fit for example, are determined using the given data. This model can then be applied to new sets of data, to place these data in context and give a better understanding of the data compared to previous data sets.

There are two basic types of models which we consider to illustrate this; classification and regression models. A classification model will be given training data divided into two or more categories; it will then compare any new data to the data in each of its categories to decide which category the new data should be placed in. A very simple example would be classifying whole numbers based on whether they are even or odd. There are numerous demonstrations of this type of model, categorising images, perhaps of dogs or cats.

Classification models can be very powerful. The ability to determine what is in a picture, or whether a data set represents a normal state of affairs or an anomaly. However, the limitations of a classification model are also apparent. A classification model will produce a binary outcome, either something is in a class, or it is not. Some models attempt to moderate this with the use of confidence percentages, but it is not an exact science. Classification models will also struggle with objects for which their training data is incomplete, resulting in well publicised examples of erroneous labelling. A one class SVM model could be used to classify vibration data as “normal” or “not normal”, but these models reach their limit at this point.

5.1.1 Regression

Regression models can, in the most simple way, be thought of like drawing a line of best fit on a scatter graph. The idea is to try and extract a pattern from many data sets. The new data sets can then be compared to these trends, and we can determine how close this new data comes to following the trend. There are various ways of ‘measuring’ the distance from the trend line, including the least squares method, or polynomial penalty method. An example of regression in a two-dimensional sense is given in figure 16. A machine learning regression model will use the measurement methods supplied to it, usually specified by the model builder, to determine trends which best match the data it is given.

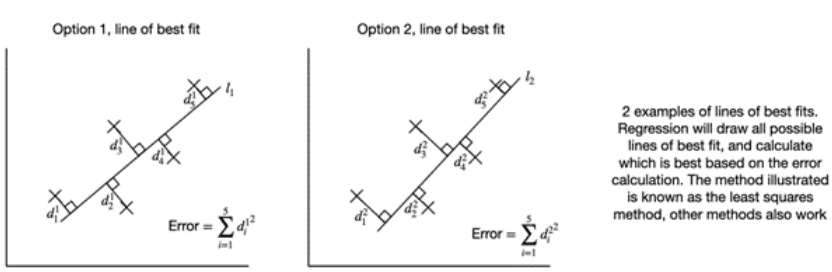


Figure 16: Demonstration of two-dimensional regression methods.

The advantage of a regression model over a classification model is that it will naturally produce a metric which will show how well the new data fits existing trends. This means we can quantify how well a system is performing. One possible disadvantage is that the trends are assumed to be correct, so a proper set of training data is imperative.

5.1.2 Autoencoder

An autoencoder is a type of neural network which is designed to learn and then perform data compression. The basic idea is to teach a model to compress and then reconstruct its own data. This will allow us to work with a simplified data set. The model uses training data to test different compression methods, different combinations of metrics, and different end features; uses these to reconstruct the original data, and measures how far the reconstruction is from the original. The method of compression and choice of features are optimised to minimise the reconstruction error.

Autoencoders are a type of regression model. Although there is no obvious equivalent to the line of best fit, they use an in-built measurement method to determine how accurate the reconstruction is. As such, they are classified as a regression model, and this is easily seen as their output, the reconstruction error, is a scalar value on a linear scale, not a Boolean value or distinct category as would be given by a classification model.

5.1.3 Stacked autoencoder

A stacked autoencoder is a type of neural network that consists of multiple layers of autoencoders. In a stacked autoencoder, the output of one autoencoder layer is used as the input to the next layer. The result is a deep neural network that can learn complex representations of the input data.

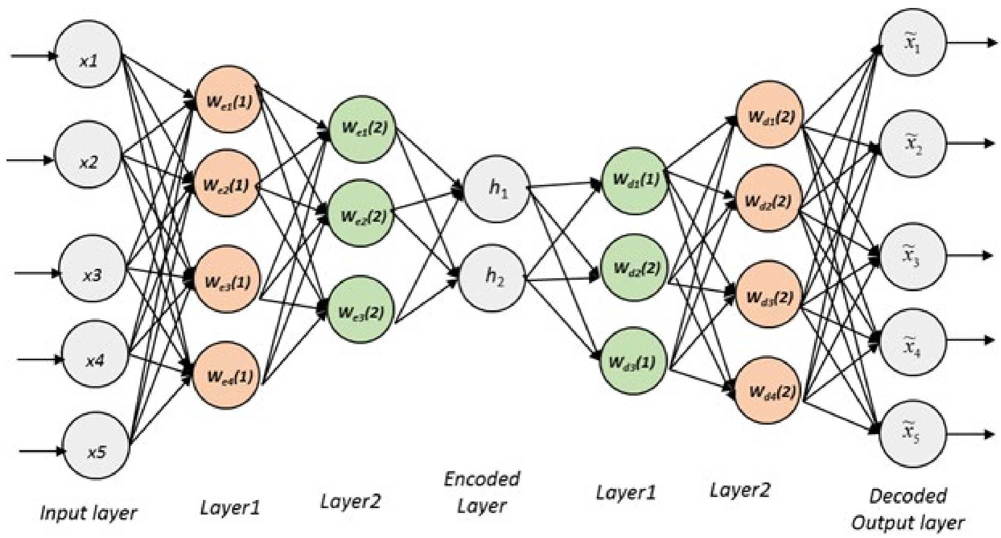


Figure 17: A stacked autoencoder model.

To use a stacked autoencoder for anomaly detection,the neural network must first be trained on data corresponding to the normal centrifuge state. During training, the stacked autoencoder learns to compress the input (feature) data into a lower-dimensional representation and then reconstructs the original input data from this representation. By stacking multiple layers, the network can learn higher-level features of the data, which may be more informative. An advantage of using a stacked autoencoder is that it can learn a more complex representation of the input data than a single-layer autoencoder. A suitable choice of number of layers in the autoencoder, number of neurons per layer, bottleneck/encoded layer size & regularisation mechanisms should be made to avoid over-fitting.

5.2 Reconstruction error

The system Dianomic is using to model the centrifuge, is a stacked autoencoder regression model. This works by taking all the calculated metrics from the data collected by both the accelerometers and from the PLC, which amount to over 1000 per WPD node, and reducing this to a set orders of magnitude smaller, which preserve as much distinct information as possible. This is illustrated in figure 18, which shows the complexity and volume of calculations being undertaken on live data within FogLAMP, with approximately 1250 features calculated.

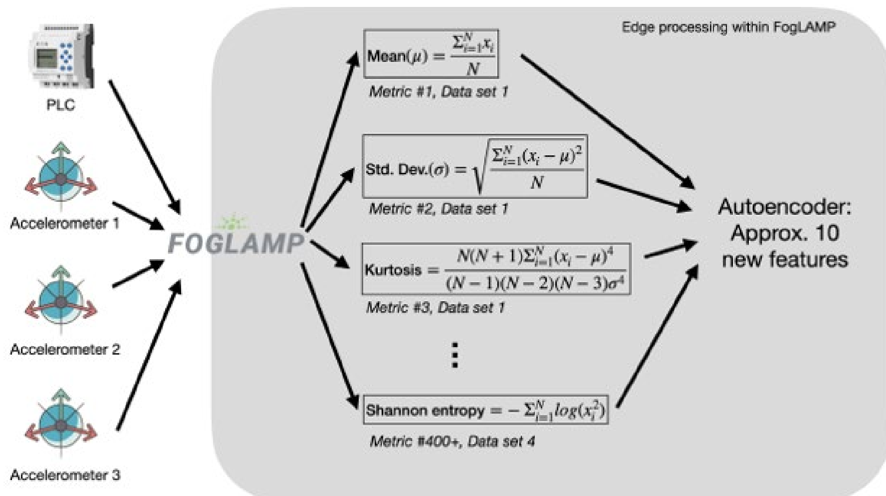


Figure 18: Diagram showing the flow of data from selected metrics through to the autoencoder

The autoencoder will then attempt to reconstruct the metrics from the smaller set of features.
The total error produced is called the reconstruction error. This reconstruction error is then used as a health metric for the centrifuge. The larger the reconstruction error, the less well the new data set matches previous data set's trends and patterns. There are several different states that the centrifuge can be in, and this can be recovered from the PLC. This will then inform the choice between different autoencoder models which are trained on different data sets.

No thresholding is done on the health metric to determine the relative health of the centrifuge, this is at present left to operators. However, in theory an automatic action or alert can be sent when the health metric exceeds a certain value. Due to the nature of the reconstruction error, this is best set by trial and improvement rather than by any closed form solution.

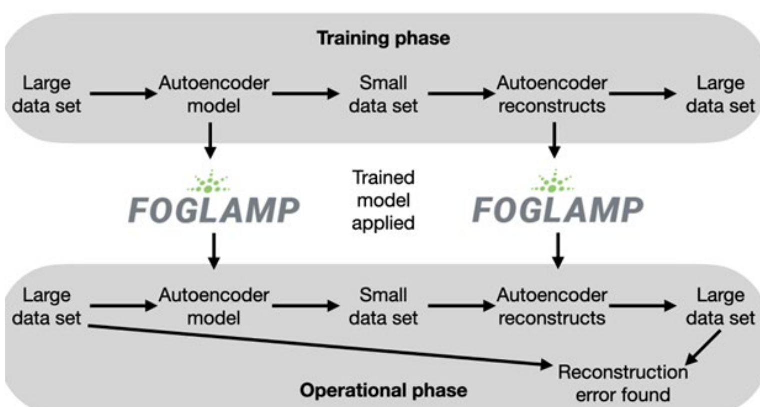


Figure 19: Diagram showing the process of using an autoencoder model to find reconstruction error.

5.3 Data sets

It is important to consider the different data sets used to create these models. Any data set collected from the machine which contains both the accelerometer information and the PLC data may be used either as a training or testing data set, or may be processed by a finished model. Any data sets which do not have all this information are incomplete and cannot be used.

At the start of operations, all data should be classified as training data, it is pointless to test or use a model which has not been properly trained. There is no upper limit on the number of data sets which are necessary for training, and more training data will produce better results, but the machine is not useful in the training phase. A good way to test if the training is well advanced is to test the model with a testing data set, note the reconstruction error, then give the model one or more extra sets of training data. If the new model produces the same or similar reconstruction error, it is well trained. This process should be repeated several times. Once the model is trained, data sets may be sent to the finished model. It is important throughout to consider the training of each of the sub-models produced by changing the machine state, a very reliable centrifuge may result in the model for a healthy machine being trained before other categories.

6 Results

The objectives of this project were to use vibration monitoring to ascertain the condition of the machines, to use this information to reduce maintenance costs and disruptions, and to maximise profits and democratise expertise. The prototype case study that has been discussed here has been extensively tested by ADM and they have found that these objectives are thoroughly met.

In their experience costs have been saved, by reducing maintenance procedures, which are themselves costly, and by reducing maintenance downtime. This has helped to maximise the profits and efficiency of their plant. Furthermore, their operatives have a better understanding of the maintenance requirements of the machine and its performance at any given time, without extensive training.

It should be remembered that we have been discussing a single case application of the technology here. The results found for a single centrifuge are encouraging, and it is clear that by using this system of instrumented machines, industrial data pipelines and edge ML, changes can be made to improve the efficiency of the machine. However, this is only the beginning; instrumentation of more of the production line, and integration of these instrumented machines will allow for faster, more accurate, and more detailed condition reporting, and root cause analysis, allowing for more efficient and more agile response.

When imagining the benefits offered by a fourth generation industrial plant such as the one we describe here, it becomes clear that continued use of data allows for ever greater refinement of the manufacturing process, more responsive and specific systems, delivering better and better results as more and more data is acquired, and more and more refinements can be made to the ML models, and more factors along the assembly line are considered. It also becomes clear that this plant is increasingly autonomous, already the PLC's controlling the machines are being used to source data, and can be easily directed to take action on results derived from the data across the plant. These are the steps to an autonomous manufacturing facility delivering greater efficiencies than even the best plants today, and at a fraction of the cost.

7 References

1. "Condition monitoring of machines: a review" by S. Tandon and R. Choudhury (2012)
2. "Vibration Based Condition Monitoring: A Review" by E. P. Carden and P. Fanning, Structural Health Monitoring 2004; 3; 355.
3. "Brief review of vibration based machine condition monitoring" by D. Miljković, HDKBR INFO, 2015
4. "Contact Mechanics" by J. R. Barber
5. "Contact Mechanics" by K. L. Johnson
6. "Forced Vibrations with combined Viscous and Coulomb Damping" by J. P. Den Hartog, Philosophical magazine 1930; 9; 59
7. "Dynamic analysis of mechanical systems with Coulomb friction" by L. Marino
8. "Elasticity" by J. R. Barber
9. "Plasticity" by A. Bertram
10. "The elastodynamic bimaterial interface under mode I and mode II loading" by B. Gurrutxaga Lerma, International Journal of Solids and Structures 2021
11. "Evaluation of numerical integration methods in elastodynamics" by G.L. Goudreau and R.L. Taylor, Computer Methods in Applied Mechanics and Engineering 1973; 2; 1
12. Information on the planetary gearbox: https://www.alfalaval.com/products/separation/ centrifugal-separators/industrial-centrifuges/px-range/px-100/
13. Figure 1: https://www.alfalaval.com/globalassets/documents/products/separation/ centrifugal-separators/disc-stack-separators/px-100-disc-stack-centrifuge.pdf
14. Figure 2: https://www.motioncontroltips.com/whats-the-best-type-of-gearbox-for-servo-applications/
15. Figure 3: https://www.quora.com/How-does-a-piezoelectric-igniter-work
16. Figure 12: https://www.continuummechanics.org/wavelets.html
17. An excellent explanation of the wavelet transform: https://www.youtube.com/watch?v=kuuUaqAjeoA
18. Figure 15: https://www.analyticsvidhya.com/blog/2021/05/shape-of-data-skewness-and-kurtosis/