Corrosion, Environmental Degradation and Surface Engineering

Prof. Harish Hirani
Department of Mechanical Engineering
Indian Institute of Technology, Delhi

Lecture – 27
Failure/ Fault Analysis

Hello and welcome to the 24th lecture of course on corrosion, environmental degradation, and surface engineering. The topic of this lecture is a failure or fault analysis. Now most of the books refer only to failure analysis, but I have added one new word here: fault analysis. Why I have added I will be explaining in this lecture. Let's revisit the first lecture where I outlined the overall framework of the course. We mentioned that it would consist of five segments. We have already covered surface degradation and its relation to the environmental factors. Additionally, we have gone over most of the important characterization techniques.

A few will also be covered, possibly when we start surface engineering-related topics. Now in the present lecture we are going to start this failure analysis, life cycle assessment, and we already have covered failure modes. We need to cover the life cycle impact assessment. Whenever we think about the life cycle assessment, we need to know what the residual life of the component is and, finally, what the economic assessment will be. Now if I want to do real maintenance on some system, will that be cost-effective, or will replacing the whole system be cost-effective?

So these points are necessary to take any decision. So for the decision-making last time we started a data-driven approach. When we have data, then we can take a good decision. We do not have data; suppose we have only models available, but that will not be sufficient to take an appropriate or dynamic decision. Most of the decisions that will be based on the model will turn out to be aesthetic decisions. It may be that those decisions are good for today, but what will happen after two years? We will not be able to estimate properly. So for that purpose we require data and then a data-driven approach to take a proper judgement, a proper decision.

In this section, we'll explore the topic over the course of 3 to 4 lectures, focusing on aspects like failure analysis and life cycle assessment. In today's lecture, I'll introduce the term 'fault' and explain its significance in detail. But first, let's review what we've already covered. We discussed various failure modes across 3 to 4 previous lectures.

Lecture 7 we also covered creep; lecture 11 we covered something like a synergetic effect of the wear, corrosion, and fracture. And if you remember the wear debris that is coming out of the surface, it should be treated as a fracture because it is getting separated from a main component. So fracture becomes a very dominating feature or dominating failure mode, and even the corrosion will weaken the material or system, but finally the wear out and the damage will happen because of wear or fracture. So the corrosion is like increasing the effect, or may be increasing the intensity or severity of the failure or severity of the fracture. That way, I am treating this the Again, we study something like a fatigue fracture. We say there are three stages.

One is a crack formation. Again, I am using the word crack formation, not a crack initiation. We know that initial cracks will be there in almost all the materials. There will be grain boundary problems; there will be some sort of additionally in the manufacturing-related problems. There will be some cracks, or there may be some lines that may not be seen easily, but there will be cracks, but crack formation, or which is really going to lead to the propagation side, we are differentiating that.

So we say the crack initiation, crack formation. Crack formation is like it has a sufficient line that is visible, and if you provide some energy, then it will be causing the failure. So crack formation, crack propagation, and failure—this is what we study. Another thing we studied was that crack initiation occurs mostly as a stress concentration point. If the components are made equally stressed, it is designed in such a manner that more or less failure can be avoided, and another, as I mentioned, is that there will be some sort of initial crack or material discontinuity that will be causing some sort of initial crack.

Another interesting point is the potential for voids or microcracks to form, which can propagate under cyclic loading. This occurs when the material experiences alternating strain—shifting from tension to zero or even to negative values. Similarly, when the stress fluctuates between positive, zero, and negative levels, repeated cycles of stress, strain, or thermal changes can lead to fatigue and eventually failure. Initially, what might start as a minor discontinuity in the material can develop into a microcrack, which then spreads and ultimately results in a fracture. We covered this in earlier lectures, along with concepts such as creep failure.

We say that creep is basically slow; it does not happen initially or immediately, the way corrosion does not happen immediately; it is a slow process. Similarly, creep is also a slow process. So slow and persistent deformation, there will be continuous strain and the continuous deformation, and we say even the load is not varying at all. There is a possibility. While we were initially thinking about a fatigue where the load continuously varies, the strain continuously varies, or it may be some other parameter like a temperature varies.

While in this case a creep, the stress is constant or may be more or less steady. So steady load, steady stress, but again we can divide in three stages. We say primary stage mostly because it is a high value. Initially, wear will be very high, then we get a steady value, and finally, up another side, that is why we use the word bathtub curve. Similar kind of happening in a creep also. We say in a primary there is a mostly initially high, but it will reduce over the time.

Then secondary, which will be more or less steady, and tertiary, which is a leading to the fracture that measures significant failure possibilities, are there. And this happens because of the excess stress, temperature, and sometimes material properties. A few materials are very sensitive to the creep failure; we should not use those kinds of materials. And one more positive point when we mention that if the material has fine grains, then chances are reduced, particularly for the creep failure. The reason being that they will not be if there is any crack, and then because there are fine grains immediately, it will get a stop at the grain boundary. It will not be able to penetrate unless there is a stress level very high.

So, this brings us to the point that fine grains can halt their expansion at interfaces, especially at void levels. As I mentioned earlier, instead of focusing solely on failure analysis, we'll shift our attention to fault analysis. Now is the time to dive into the details of that approach. The overall aim of this course is to detect failures before they lead to breakdowns, excessive wear, economic losses, or environmental degradation—all of which we want to avoid. To achieve this, we need to develop a sensing strategy—whether time-based or frequency-based—and understand the nomenclature or signature associated with different parts and features, as each will exhibit unique behaviors.

So, the sensing trends or frequencies that are specific to parts, their function, their design, and stages of the wear, like the ultra-mild wear will have some sort of characterization, and the and the medium wear will have some different characterizations. So, if we go ahead with the training analysis, the time-based or frequency analysis, we will be able to dig out some sort of key points that need to be examined, which need to be recorded and plotted, and then we do need the needful. So, that is what we are using when we record a data-driven approach. Now in this case, the signal can be a time domain signal, a frequency domain signal, or a combination of the time and frequency domains. So, depending on the situation, which signal will be more important, we can go ahead.

So, this is to failure modes, as I say, different kinds of failures. What is really our aim to detect the failure well in advance and then stop it if possible? And then what is really required for that purpose? We required some sort of sensing unit, sensing equipment, or sensors that can really give different kinds of signals, like a time signal, a frequency signal, or a time frequency signal. Some sort of algorithm will be helpful to help us come to the decision. And data-driven approach, we really required good data, and then based on that, maybe some good algorithms that can really provide good results. So, last time in our previous lecture we discussed something like a PCA, which even though we have 10 features initially, we extract only 3 features or 4 features that are dominating and playing a major role, maybe 85 percent (85%), 90 percent (90%), then I will not really think about remaining 7. So, those kinds of data cleaning and then dimension reduction are important when we think about the failure mode analysis, or maybe the failure mode and detection of the failure and remedies of those methods.

Now when we discussed earlier failure modes, we said that major failures we showed as a fatigue failure and a creep failure that has been repeated here. In this scenario, the high stress location is typically initiated by a small fracture. So, there will be the possibility of a notchhole or surface defect, and this is a major thing if we get something more like a fatigue failure. So, we need to really see if there is any notch, or maybe we are making any hole, or any surface defect or surface damage, or maybe some surface features that are required for some other purpose but will enhance the fatigue failure. We know that the S-N curve can be utilised particularly for the ferrous material, which has a basically infinite life (number of rotations is more than 10^6).

In this context, when the lifespan exceeds 10^6 cycles, we refer to it as high-cycle fatigue. Conversely, if the lifespan is significantly shorter, we term it low-cycle fatigue, where failure occurs at that level. In the previous slide, I discussed primary, secondary, and tertiary creep. As you can see, during primary creep, there's a sharp change in the slope, indicating a rapid transition to a steady-state phase, which is represented by a more linear curve. In contrast, when we talk about rupture—rather than fracture—there’s a sharp increase in the slope. These are key points to consider.

Now, why do we need a fatigue S-N curve? It helps us determine the fatigue strength, which is crucial for predicting where failure might occur. Through experimentation, we can identify the fatigue strength, knowing that failure happens due to the continuous reduction in the cross-sectional area caused by cracks, expansion, or necking under tensile stress. Initially, pores or cracks may expand until they reach a critical point where the effective area becomes too small to sustain the load. At this stage, the material will fail, either through brittle fracture or ductile failure, depending on factors like the material type, temperature, and other conditions. These concepts have already been covered in the course, including discussions on cleavage, separation, and grain boundary separation.

So, those things were discussed, and then I am just picking up this slide or maybe these figures from the ASM handbook because they have a complete handbook on the failure analysis. And basically what we are talking about is a separation or the decohesion, and that is a rupture. You can see here the example of the landing gear that failed because of the stress corrosion cracking, which was also discussed. So, stress corrosion cracking is one of the failure modes we discuss in detail. Now what we are able to see here is that because of the decohesion, there is a separation in the grains themselves.

So, these are basically the planes or the lines where the failures will start in a big way. Now if you look at it much in a magnified way, we are able to see that the failures are happening because of some reason. Now corrosion was one of the reasons, and we also discussed something like a hydrogen embrittlement that is shown in this figure tube, which is a zoomed area of this. So, what we showed in this is that this material is AISI 8740 steel, and in this in the nut formation, this is basically a nut, and the failure of that has been discussed. What we are saying in this case, particularly when the cadmium plating process is started and then sufficient, and then the backing was not done because of that, whatever the hydrogen generated in the process, that remains there, and what we say is the retaining hydrogen itself.

Now we know very well that hydrogen will retaining will cause a diffusion or maybe the spread near the grain boundaries and weaken the grain boundaries. Not only weaken the grain boundaries, it can even cause an intergranular fracture. So, these things we have already studied in the earlier case. Now the question comes: these are the kinds of things we see after the failure, where the failure had happened, we can study, we can provide a remedy after that. So, in the near future, we can remove, we can modify the process, we do a sufficient backing, maybe give a little more time, or maybe there is some sort of when a stress corrosion is happening, we avoid corrosion as much as possible. So, those things are there, but it is after the failure; it is not before that, and what we are talking about in this domain of failure analysis is the fault analysis we say before real failure occurs. Can we act on that?

What we truly need is a sensing system, as I mentioned in the previous slide. We require sensors to gather crucial information, such as the presence of cracks. We've discussed various techniques for deriving this information, where signals like vibrations are measured and other parameters are inferred from them. These are often indirect relationships. Our focus is on both direct and indirect sensing methods. To establish these, we rely on extensive literature reviews to build principles and models. However, even after developing these models, they may not work perfectly in all situations. This is where the data-driven approach comes in. We start with established models and literature, integrating them into our sensing or computational units. As new data becomes available, we continuously update and refine these models, improving their accuracy over time.

Under varying environmental conditions, the behavior of the material and the system will vary, despite the interdependence of the parameters. So, what will be the final result as such may not be the same in every situation. So, that is why we really required a data-driven approach from that point of view. Now here we the three diagrams also shown, which crack has been shown, which splitting has been shown very clearly, and then finally, here we are also saying that there is a cavitation erosion or cavitation failure, which are also visible in this. So, the failure modes are many failure types, and many questions come. Are we able to convert to the digital form, are we able to really come up with something that can be made available to the computer, or can we really make an automatic system for that purpose?

So, the challenge lies in transitioning from offline to online systems, a process that is currently underway. Traditionally, we detect failures using offline methods, such as SEM or other advanced microscopes, after a failure has occurred. These analyses are conducted in an offline mode. However, the current focus is on whether we can shift to an online system that has the intelligence to make decisions and judgments as needed in real time. This is the direction we need to move toward.

Another interesting observation that I've included in this lecture is related to indentation tests. When we conduct these tests to determine the hardness of a material, it's crucial to consider the specific requirements. The indentation test is fundamentally about assessing material hardness, which is a highly important aspect.

In this case, if we proceed with applying some load using an indenter—where the hardness of the indenter is significantly higher than the material being tested—we leave it in place for a certain period. The authors of a 2009 paper, which I'm referencing, discovered a phenomenon called non-indentation creep. They found that by using nano-indentation, it's possible to also assess creep behavior. This demonstrates the ability to derive multiple parameters from a single test. For example, they observed that keeping the indenter under load for just 0.01 seconds revealed a distinct mark, indicated by a star in the data.

Now, if instead of the 0.01 second if they hold for the 50 seconds, there is a growth, and growth is 2.6 times the star mark will be much bigger in a size, and that is what we indicate that there is a creep behaviour that can be derived from this. So, not necessary every time we need to come up with a new tool as or maybe the new sensor; we can relate with the existing sensor in a much better manner, and then we can drive the relation again for that purpose. A data driven approach will be better, or maybe more useful. So, in this case, the size of the indentation found in the creep test they use a 10 milli Newton load, and then the contact edges. You can see the contact edge here; this is the contact edge; this is the contact edge in star form. So, this is the three-point star, and then as the creep time increases, as the creep time increases, this spread will be on a higher side, or it will get bigger and bigger.

So, that is what we can do, and in fact, more studies can be done in this manner. Now, another thing I want to just refer to, the ASM handbook has published a number of atlas for the different, different materials. As a result, I am unable to cover all of the materials for every type of failure. So, that is why I am giving this as a reference to what we call an atlas and then an atlas of the microscopic image. We aim to illustrate the copper alloy and then demonstrate the behaviour of copper at various temperatures.

In this case, the first scenario, labeled ABC, was conducted at a temperature of 425°C with a tensile strength of just 20 MPa, which isn't a particularly high stress level. Here, temperature plays a significant role. As shown, the sections labeled A and B are magnified, revealing very smooth surfaces—all indicative of failures. Despite the low stress of 20 MPa, failure was observed after 76 hours and 54 minutes, with B and C marking the points of failure.

In another test condition, they increased the temperature slightly by 15°C (from 425°C to 440°C) and simultaneously reduced both the temperature and stress by 25%. The result was a notable change—the failure occurred after 308 hours and 30 minutes, almost half the time of the first scenario.

So, even with increasing temperature and decreasing stress, the failure time has increased by 308 hours. Now, this 308 hours is almost 3 to almost 4 times of this. So, 4 times compared to this. Now, what is really happening? Again, we truly need this information; it is published in the ASM handbook, where we can view images of the various materials.

You can observe the sensitivity here. By changing the temperature by just 15°C and reducing the stress by 25%—from 20 MPa to 15.5 MPa—we see a significant increase in the material's lifespan. This indicates that predicting residual life requires more complex modeling, as simple changes in parameters like temperature or stress levels don't always yield quick or accurate results. This is why a data-driven approach is necessary for this type of work. Additionally, in this case, it's worth noting that the grain facets are oriented at a higher angle relative to the tensile stress.

In this case, the small dimples visible in the zoomed image are typically caused by micro voids present in the material. As I mentioned earlier, every material contains some defects, including micro voids. When these micro voids come together, especially under load or high temperature, they merge to form larger voids, which appear as dimples on the surface.

Now, turning to the main topic of failure, we define failure as the point at which a system or component no longer performs its intended function. As discussed in an earlier lecture, a system is considered to have failed when it stops working entirely, or when it no longer serves its purpose effectively. For instance, a laptop that's 15 years old might be considered a failure if it has lost its usefulness, even if it still powers on.

Today's results are different, and perhaps the processing speed is much faster. We are using an I9 processor, and maybe that time it is an I3 and it turns out to be useless in that way that also the column will cause us a failure. However, we are not treating that way, and then another one is that performance drops significantly and complete breakdown. So, watching and observing this phenomenon is not difficult. So, I do not really require very sophisticated analysis for the failures usefulness has come down. I can be rejected, and then when the when the question comes, performance in drop can be rejected.

Detecting failure is not particularly difficult. However, identifying the root cause of the failure requires a more complex approach. To uncover the root cause, we need to delve into a multi-layered analysis, starting with a simple fault and observing how it progresses, or how multiple faults may combine to create a problem.
In my view, it's crucial to analyze faults. Failures are already apparent, so our task is to trace them back to their origin and understand how they developed from that point. Failure detection itself is straightforward because it becomes evident when something isn't functioning properly. The real challenge lies in fault analysis. Properly identifying faults is essential, as it makes detecting failures much easier.

Some people refer to failure as a system crash. While identifying the crash is usually not difficult, the analysis can be more challenging because we may not immediately understand the root cause or where the issue began. And then we really required a couple of approaches to really detect that. Another time people use the word failure where it is not in serving the intended function, which is the same thing that has been mentioned over here. Now, what is the best control strategy to avoid the failure? We say identify the possibility of the failure as soon as possible. When the system is getting loaded, or maybe when a system is starting to perform, if you are able to think about the possibility of failure, then it will be easier to understand. That means it really requires sophisticated many features that need to be addressed and need to be really continuously seen that way.

Now, detection of the failure is easy, as I mentioned, but what will be the probability of the failure? Now, if I have a database, I can see one failure has happened in 308 hours, another failure has happened after 500 hours, and the third failure has happened after 1000 hours. I can daily use a statistical method to figure out what would be the probability of the failure if I go with a new system, or maybe we use again data to come up with the results. So, this is why we required really historical data; we really required failure data to do analysis, but again, that will be completely mathematics. We may get exactly if there is even 5 percent (5%) of the failure, and the probability of failure we may get 5 percent next time, or we get only 1 percent 1%, or we get 10 percent (10%). So, we really require more sophisticated analysis. For that purpose we required really key factors or key performance indicators, or what we use our KPI, and then some sort of inspection procedures also. Where there should be an entity will be based on the sensor and what kind of sensor we are using; those things are important for us.

If we adopt the mindset of implementing the best control strategy, we should aim to receive information and take corrective actions before any failure occurs or is imminent. This is why the traditional corrective maintenance strategy should be replaced with a predictive maintenance strategy. We'll discuss this in more detail after two lectures when we explore different maintenance strategies, including predictive maintenance, and review some case studies.

We began by discussing failure, then moved on to faults and fault detection. As I mentioned, detecting failure is not particularly difficult; the challenge lies in detecting faults. If data on the probability of failures is available, it can serve as a starting point, or we can use a model as the initial input for our work. From there, we need to focus on key performance indicators (KPIs). If we're unsure which KPIs are critical, we might start by measuring 100 different parameters and refining our approach through iterative analysis.

Initially, we might hypothesize that factors like vibration, acoustics, or eddy current response will play significant roles. While brainstorming, we can consider many potential variables, but it's crucial to identify the key parameters. As discussed in a previous lecture on Principal Component Analysis (PCA), we need to determine which components are most important—those with high eigenvalues and significant impact.

So, those things are important. Now as I mentioned about the failure, we can then another thing comes: can I categorise the failure? So, we're talking about the categories; can I say that there's a low chance of failure? The probability of failure can be classified as low, medium, or high. If I do not have much knowledge, at least I get the rudimentary level or the initial level. I can start with a low chance of failure, medium chances of failure, or high chance of failure, and then it is like a fuzzy logic, where we do not have a clear idea but at least some vague idea.We can first categorize, and then proceed to a more in-depth level. So, fuzzy logic-related parameters are also there, and they can be given initially.

We will discuss these concepts further in the next lecture, particularly how recurrence time can be used to develop an understanding of failure probabilities. Now, what is considered a medium failure, and how do we categorize it? A medium failure occurs when a material or system lasts as long as expected. For example, if a system is expected to last three years and it fails around that time, we categorize it as a medium failure—not a low failure, which would mean it consistently lasts less than three years. On the other hand, if a system with a three-year expected lifespan fails after just a month, we consider it highly critical.

This leads to an increase in the severity index. We will use these terminologies in our next lecture, where we will also discuss Failure Mode and Effects Analysis (FMEA). FMEA is a systematic method used to identify potential problems in processes, whether in manufacturing, assembly, or any other system. It applies universally, and we use it to determine the types of failures that can occur and how severe those failures might be. As previously discussed, we need to examine the causes of these failure modes and the issues that lead to them.
By conducting this analysis, we can prioritize the areas that need improvement. We rely on historical data for this analysis, which helps us identify which areas are most critical and require more focused attention. Now that we've covered failure, let's move on to the concept of faults.

We see in engineering a system reliability word fault means mistake in a design and any flaw which is made in either in a manufacturing time or may be some other time, and then we may not be knowing that kind of fault that is something like a hidden fault something or something that is not visible to us, but it will cause a finally failure, and that is why there may be even the 5 percent (5%) failures and may be happening because of the faults or hidden faults. If you analyze that and find out that 95 percent (95%) of systems are successes and only 5 percent (5%) are failures, quite possibly there are some flaws in a design. There could be flaws in the manufacturing process, in the assembly process, in an operation, or there could be a human factor at play. So, those things are related to fault, which is not related to failure immediately. So, as we say, any imperfection is also important in that case. Now coming to the imperfection point, or may be a flaw point or may be a mistake point, we need to really treat this as a fault.

Now, faults could be causing the failure, but as I mentioned, they will not appear immediately. So, immediately we will not be, may be after a month, may be after some duration it will come. So, I am just taking one example, something like corrosion in a brake line. The corrosion is happening often because it is open to environment brake lines, and then corrosion will occur there. Now if this failure occurs, then naturally there is a need to really look at if the corrosion is happening and what will be the life cycle of that component or system. So, that is why the fault and then detection of fault, and because of that fault, what will be the impact on the life cycle? That is important, and that is why it needs to be assessed properly and then managed properly.

Now, as I mentioned very clearly, a fault is a hidden or latent problem in a system; it is a hidden problem, and we need to diagnose. We have already discussed NDE techniques. We know that on the surface, if there is any fault, it can be seen, but if there is some fault either within a surface or subsurface, then we have a number of techniques. So, those techniques depend on the number of sensors. Can we utilise those sensors in our way and come up with some new technique? And then we say fault diagnosis, and that is how many times the conditional monitoring will be discussed, as I say next to the next lecture. Next, another example. As I mentioned here, corrosion in a brake line could cause a line to break, particularly under high pressure when the area of the fluid is decreasing. Naturally, as the pressure increases, the inside of the brake line begins to corrode. Even what you say, the brake will not be visible, or even the pipe corrosion will not be visible, and the area's continuously decreasing pressure will build up on that.

And then, which is not visible, one day what will happen is the brake, the whole line will burst and act as an instantaneous or catastrophic failure, and then in that situation the brake will stop. So, brake will stop means brake will fail, but if there is a fault initially that was not visible, that will be a kind of failure. What is happening is somebody is driving in the car at very high speed, and suddenly the failure of the brake happens naturally. It will be very dangerous for not only the person who is driving but also for other people around them. So, this is important, and another is mentioned finally: identifying the fault and fixing it. The question comes: how do we fix it? So, there are a number of methods also to fix it, as we will be learning a few methods to fix those faults in surface engineering methods. So, that is possible, and the system becomes more reliable if we are able to fix it. If we are able to identify faults and fix them, then the system can be more reliable, safer, and better, and that is our aim, of course, to make the system more reliable, safer, and better by utilising and understanding the failures and faults and then doing the remedial action and mitigation action.

The goal is to develop a reliable, safer, and more efficient system that benefits society while minimizing environmental degradation, creating an overall win-win situation. To achieve this, we need several key components: modeling, signal capturing and processing, and computational intelligence. These are essential elements of a data-driven approach.

Modeling is particularly important because it provides a basis for comparison—new data must be compared against a reference, which is where modeling becomes valuable. This modeling is built on previous data and knowledge. Next, the signals captured by sensors need to be processed accurately. By applying computational techniques, such as Principal Component Analysis (PCA) or other artificial intelligence algorithms, we can derive useful insights from this data.

In essence, when any feature or parameter of a system deviates from its expected value, it indicates a potential issue. The expected value is determined by past experiences, historical data, or information gleaned from digitized graphs. This approach helps us predict and address problems before they escalate.

So, expected value will come from that. So, material is supposed to give a 300 MPa, and that has naturally been verified, and the data have been really available in the data book, which is why we are getting. So, that is the expected value, and then when we use sensors and find any deviation from the expected value, we say there is a chance of the fault. Now, we need to consider the possibility of noise, which is why we use the term "signal processing." So, fault, and then if we do signal processing, we remove all the noise, and then still we find a deviation compared to the expected value, we will say there is a chance of the fault. So, diagnosing faults using this kind of sophisticated device, or maybe sensors, processor, and then the right algorithm will be able to identify the faults, and that is what we really require, and that is the main tool now in maintenance engineering.

So, that will be done. Of course, once we identify the fault, it will also assist us in determining the tendency towards degradation. If the fault is increasing very fast, or not increasing at all, it has halted. I use an example: if you go for the very fine grain boundaries or grain structure, the possibility is that the fault started and immediately got halted; it is not really able to penetrate the other grain. So, it is along the surface of the grain or along the boundary of the grain; it is not really increasing the fault as such. So, there is a possibility, and the overall aim of this course is basically to proceed with good fault identification and removal. So, that is why the main goal of failure analysis is to proceed with fault detection and diagnostics.

That's why, to put it simply, the term FDD is frequently used. So, we will be going with the fault detection; we are not going with a failure detection; we are not going with a failure analysis; failure analysis will happen. If we are having a reliable FDD, then failure analysis will come its own way. So, this is important, and then sometimes people, instead of the FDD, use a word-of-fault diagnosis and health monitoring. That is why next to next lecture we will be giving on the condition of monitoring the units, and it will be highly related to each other. Now, the FDD approach, as I mentioned here, requires modelling, signal processing, and computational intelligence.

So, the FDD approach can be a model-based online data-driven technique, or it can be knowledge-based. So, initially, we can start directly if we have historical data; we can go ahead with the knowledge-based data, some sort of model, which is why the modelling is required. However, if I do not have anything and I make a totally new thing, then quite possible initially the sample will be made and some data will be generated, and based on the data we can really make an approximate model, and then we continuously refine that model. So, that is also possible. So, our main aim in this section is to go ahead with the FDD, something like fault detection and diagnostics, and then possibly give a remedy using the surface engineering knowledge.

So, this is what we know we say: monitoring the faults How do we go ahead with monitoring the faults? It is analysing the fault or multifunction, as I say; it is not really deviating from expected performance fault level, not a failure level as such. In a system, a process can be made easier by using the data-driven approach that was described. Now, we would be really required initially, as I said, if we have data, it will always be helpful. So, include past data so that we can draw some sort of conclusion based on the relationship.

It can now be regression analysis or some other analysis. So, that we really required relationship between the different inputs and specific value occurrence. So, most of the NDT techniques are based on this; they have some sort of relationship established, and then whatever they are measuring, they are able to drive some other form, which may be a result, and they are able to drive the results based on that. Now, we say that gathering relevant information about the analysis must be a starting point. So, the literature review, literature documentation, and then initial model that are available in the literature must be understood and must be utilized. Now, the regression analysis and then correlation we studied last time may be employed along with a more modern approach, something like clustering and classification.

So, what is the classification? I'm aware of five distinct types of failures. So, what is the really appropriate KPI for that in the failures: failure 1, failure 2, failure 3, failure 4, failure 5. Now, 6 failures if I get, I try to give in one of the kinds of buckets: failure 1 bucket, failure 2 bucket, failure 3 bucket. So, this is basically to classify in the different kind of, like, you know, there is a crack. 1 bucket or crack is one class, or then we say that pit is another class, or we say dimple is another class.

To proceed, we should incorporate any measurable KPI-related factors. These factors are recent and necessitate the use of machine learning algorithms for clustering and classification. While we won't be covering these topics in-depth in the current course, we will highlight a few features and possibly explore one case study.
When we refer to sensor data, operational parameters, maintenance records, and environmental factors, all of these should be considered. For effective fault analysis, it is essential to review all these elements, as relying solely on sensor data is insufficient. We must also account for operational parameters to establish correlations between sensor readings and the actual operational conditions.

Additionally, maintenance history, such as the timing and duration of maintenance activities (e.g., after 3 or 6 years), is important. Lastly, environmental conditions, such as humidity and temperature, should also be factored in.

This is why there is a variation in the performance of the different materials, and one system works well in one situation but not well in another. So, those things can be dig out using this kind of data, and that is very important for a data-driven approach; we really required a data-driven approach. Now, what we mentioned is another one that the pre-processing of the data is required. So, when you do the preprocessing, mostly we try to denoise it; we remove the noise, and then sometime, if some data have not been recorded, we do some sort of interpretation also. So, the missing values will be there, and somewhere we get some random out on the very high value kind of thing.

So, remove outliers also, and then every variable in the last PCA also mentioned that we need to go ahead with the standardisation because we are not going to deal with only one variable. We are dealing with many variables; whenever there are many variables, standardisation is a must. And then if there are any constructive features coming out of that, or maybe the new feature that we need to be diagnosed, that needs to be highlighted. Now, once a model is being made, we do some sort of training to model. So, that it is really meeting the expectation, giving good results, and confirming the predictive analysis can be performed.

In a predictive analysis in the future, we are trying to predict what will be happening tomorrow after that, and the whole predictive maintenance is based on this kind of principle. We can now say that if we keep getting more and more natural information, we will refine our model. So, naturally, we need to have an algorithm that can continuously refine the model and improve the ability to foresee or explain the potential failures. So, this is the main thing: if we are able to monitor the faults, if you are able to utilise a data-driven approach, and we are able to make a predictive model, and we are able to continuously improve that model, we will have substantial benefits. We can really minimise the surface degradation, environmental degradation, and even the corrosion, but for that purpose we really require this kind of approach, which continuously gets updated because everything is iterative.

If we do not update whatever we design today, we will not work tomorrow. Therefore, we must evaluate the case tomorrow and subsequently update it. So, what we see when looking for ways to improve a system, procedure, or product design—whatever I mentioned, fault analysis must be utilised, and it is itself an iterative procedure, and as a newer and newer thing keeps coming, we need to incorporate it into our model. So, that is why we say data-driven approach to the fault analysis, and then whatever we are learning from some sort of fault failure, or and then the failures can be useful for improving the reliability, quality, and performance. So, this is if we are able to really make up or we are able to really go ahead with a data-driven approach and then finding the faults, and then go ahead with the surface engineering to really bridge those gaps and minimise the faults. Let us take a one-case study. What we are trying to say is that fault monitoring, as I mentioned, is important for us.

We are exploring fault monitoring using Principal Component Analysis (PCA), which was covered in a previous lecture. This case study is drawn from a 2003 paper that applied PCA to extract features and classify gear faults. I will present this case study in the following slides.
In this study, gear conditions are categorized as either healthy (normal operating condition) or faulty. Faulty conditions include subsurface cracks, which are not visible to the naked eye, and broken teeth, which are detectable by visual inspection. This analysis aims to differentiate between gear failure and faults, potentially aiding in fault diagnosis.

So, faulty gear can go ahead with the failure, right? Now, what diagnostic information for categorising the gear failure using principal PCA? They try to use a vibrational signal. As I say, we really required some input. So, they use a vibrational signal using an accelerometer, and then they perform experiments, and that too on an industrial gearbox, and then in the hope that PCA will be able to differentiate gear faults. So, this is what they use a PCA to differentiate gear faults: if normal gear is a case 1, maybe say a cracked tooth has a 2 and the broken tooth has a 3.

There are three conditions being analyzed: normal, cracked tooth, and broken tooth. The diagram illustrates a complete gearbox, where z represents the number of teeth on the gear, and f denotes the frequency at which the gear teeth come into contact. This frequency indicates how often the gear teeth meet.

For the analysis, two types of parameters were chosen: temporal and frequency-based. The accelerometer data was collected and pre-processed, followed by feature selection. This selection process then led to fault classification. In this case, there are only three types of faults to identify: normal gear tooth, cracked tooth, and broken tooth. The analysis focuses on fault detection and diagnosis, aiming to determine the presence and origin of faults.

So, for that purpose, we really required another feature, like in this case a time-dependent feature, the temporal feature; another one is a frequency-based feature, which has been shown. Now, if I go with only the time dependent, we have a mean value as a time, a standard deviation time dependent, skewness, kurtosis, and the time dependent statistical features. So, this is what they tried to calculate the 10 parameters, and then they captured the raw data, as I mentioned, the vibration data, and those data with the amplitude versus sample point have been shown over here. Now, you can see here that in this figure the vibration level is relatively low, while in this case some places vibration is increased right.

Notice that the vertical scale ranges from 0 to 500 and -500, which are the same. However, for the broken tooth condition, the plotted range is 0 to 1000 instead of 0 to 500, and -1000 instead of -500. This adjustment reveals distinct peaks occurring at regular intervals, indicating some characteristic frequency. They performed online measurements to gather all the data under three operating conditions: normal, cracked tooth, and broken tooth. They also utilized non-dimensional factors and parameters for their analysis.

We know then PCA we required a standardisation even though these are all the factors; they are non-dimensional. Non-dimensionality is one thing, but standardization is another. So, we really required a standardisation if we are getting the vibration data. So, first, we try to calculate what the KPIs are that we have used. In this case mean value, standard deviation, HQNS, kurtosis, max peak value—that is what has been mentioned—and then RMS value, crest factor, shape factor, impulse factor, and clearance factor—and then these values they try to figure out. They collected 23 data sets for each case: 23 data for the normal case, 23 data for the cracked tooth, and 23 data for the broken tooth, and the only thing they kept was a sampling frequency. The operating speed was the same and maintained, which is why they kept a sampling frequency the same.

The issue they encountered was that the signals were overlapping significantly, making it difficult to distinguish between them visually. Even when they attempted to extract features, they found considerable overlap, which limited the effectiveness of their analysis. Since they did not provide the raw data, we took a different approach. We digitized the graphs they had plotted and used MATLAB coding to extract the kurtosis values from these digitized graphs.

As you can see, each of the nine plots consists of 10 features. So, feature 1 is a standard deviation, HQNS, kurtosis value, maximum peak value, mean value, and everything has been written in the MATLAB code. And then we can see here that this data is basically for normal gear features for the cracked feature or broken gear tooth. So, these have been vertically plotted, and then occasionally we are able to get a very closeness, something like, you know, normal gear and a cracked one. You can see the features are almost merging in this case. While in this case particularly there is a separation here, green and red are matching, and what is the green? Green is basically broken teeth, red is for the cracked one, and blue is for the normal one, and that has also been mentioned here, the normal cracked one and the blue one. Now, we see most of the time the normal and cracked ones are really superimposing each other, except in this case kurtosis value, the normal gear value has a very high.

For the most part, the graphs show that the normal and cracked conditions are almost identical, with no significant differences that would indicate a distinct feature variation. However, in the case of the green line, there is a noticeable separation in almost all instances, except in certain areas such as the kurtosis value and skewness, where there is considerable overlap.

Next, we applied standardization, as discussed in the previous lecture. When standardizing, we typically see a range of values like 115 to 165, -1.5 to 1, 0 to 30, 200 to 1400, or 70 to 130. However, after standardization, the values usually fall within a range of -3 to 3. As a result, many of the curves become almost indistinguishable, with significant overlap. This superimposing effect makes it difficult to differentiate between features. Nonetheless, standardization remains crucial for ensuring the mean value is 0 and the variance is 1.

Therefore, we can accurately determine which parameter is significantly influencing the data. Now, here without PCA we are not able to figure out, but you can see the vertical axis by and large vertical axis is same and horizontal, of course, we have 23 points. So, horizontal is same, and vertical is also same. Now, we need to use a PCA to separate the importance, which is a more important, or maybe use a PCA or use a and the eigenvalue to figure out the eigenvalues, and then go ahead with a descending order. So, in this case, first, to go ahead with the PCA, the first thing is to go ahead with the covariance matrix, and then the covariance matrix is basically relation among the members or maybe the features.

What is the relationship between the first feature and other features? For instance, the correlation between the second and first features is 0.19, and between the first and third features, it might be 0.31. Now, consider the relationship between the first and fifth features, which is 0.95. It's important to remember that the diagonal elements of the covariance matrix represent variance, and this variance should be equal to 1.
As I mentioned earlier, after standardization, the mean value should be 0, and the variance should be 1. When looking at the covariance matrix, the diagonal elements should also be 1. These are important cross-checks, and MATLAB code has been used to verify this. The diagonal terms in the covariance matrix—such as the relationship between feature 1 and itself, feature 2 and itself, and so on—should all equal 1. This indicates a 100% match between a feature and itself.

So, it cannot be other than 1. Now, when we go ahead with PCA and then we try to arrange in the descending order. So, the first eigenvalue comes at the at the top, which is a 0.3, and the and the last value is a 0.02. You can see here that more than the 0.51 is something like 5 features, and the remaining 5 are not really useful as such. So, one way is that we can really simply reject these 5 eigenvalues, and we say principal component of 5, or sometimes people say no lesser than 1 percent. I do not really care. So, we can go ahead with the d1, d2, d3 also. So, remaining all other parameters, eigenvalues are less than 1; it can be.

However, often we go ahead with the percentage base. That's why we've represented it as a percentage. The first eigenvector shows a percentage of 54.6, the second eigenvector shows 18.02, the third eigenvalue is 10.31, the fourth is 8.69, and the fifth one is 5.62. So, what we say is that 5 percent (5%) either you go ahead with more than 5 percent or total summation should be more than 95 percent (95%) . So, this is a total summation of the 5 features that turns out to be 96.88 percent (96.88%). However, it is subjective; people can think about 85 percent, 90 percent (85% - 90%).

In this case, we have chosen the top 5 eigenvalues, and based on that, we can try to recast the data. When we go for the recasting of the data and then again plotting, we can see here very clearly that after plotting, the first eigenvalue related to the to the first parameter, second parameter or second feature, third feature, fourth feature, and fifth feature. So, here, everywhere we can see the clear differentiation between the green line and the two lines, the green line and the other two lines. So, that is to indicate very clearly that the broken tooth signatures are very clear. We are able to diagnose very clearly without any problem.

If we go back to the previous one here, we say that except skewness and skewness and kurtosis, we are able to differentiate. And of course, in this RMS value also we find some sort of superimposition while in the in the new 5 features we are able to see clear division. Now, when we compare this cracked one to the normal one, we can see some sort of separation. So, future 1 we are able to see separation; feature 2 we are able to see separation; feature 3 we are able to see separation; and feature 5 we are able to see separation. While coming to the 3 and 4th features, we do not find that kind of a variation. Now, I can consider two possible scenarios: either there is a subsurface crack present, or the vibration signal is not strong enough to detect that crack.

And we need to change the sensor, or maybe we get another kind of sensor to get a better response, or maybe we go ahead with the next stage of the processing. So, PCA is one technique that we have used, but we should go for some other techniques so that we can filter out those things. These techniques hold significant importance, and we should consider them. Now, when we plot on a 2 axis or 3 axis, these are the plots here; you can see here the somewhat separation; blue colour lines are visible; there is a separation here; green colour lines are separation.

So, gear faults using only two principal components; gear faults using three components or principal components. While in 3D plots we do not find much advantage we are not able to really differentiate, but in 2D plots we are able to see here there is a separation of 2D of the broken teeth, the normal teeth and then the cracked teeth. Now, this illustrates the need for additional processing on this type of signal, as we can distinguish between broken teeth and normal teeth. However, some argue that a crack is a fault and requires proper diagnosis, which cannot be achieved with this type of signature, vibration analysis, or PCA alone. We require a little more sophisticated method to go ahead or maybe get better results.

Once the question comes, can we go with a multiple-time PCA? So, that does not really help, and there is a first PCA that has been done. We go for the second PCA, and we do not find much difference in the principal component. So, it should be rejected. We say the sequential PCA cannot be utilised; we really require some other method to come up with better and better results. So, we will continue this topic in a next lecture, and mostly it will be the failure mode and effect analysis. We will highlight that we will cover two case studies in that lecture. Thank you for your attention.