ÉCOLE CENTRALE LYON

HEXAGON MANUFACTURING
INTELLIGENCE
REPORT

Data Science for Electric Battery Performance

Student :
Neirouz BOUCHAIRA

Supervisor :
Clothilde MINFRAY
Internship Supervisor:
Moncef SALMI

August 2023

Abstract :

This internship report describes a project within the Materials Center of Excellence (CoE) that focused on using machine learning algorithms to simulate battery performance. The internship was aimed at battery innovation to uncover new business opportunities for Hexagon. This was the exploratory phase where research was based on the databases found. The report summarizes the approach, the challenges encountered, the results of the simulations and their potential implications.

The results of the incubation studies indicate the viability of implementing AI and machine learning techniques for battery performance research for Hexagon. The simulations carried out throughout the internship hold promise as a cost-effective and efficient means of testing innovative concepts for advanced battery materials.

Contents

1   Internship Context

1.1   The Company

Hexagon Manufacturing Intelligence

Hexagon is an expert in digital reality solutions, combining sensors, software and autonomous technologies. Its solutions enable customers to harness rapidly growing amounts of data and put them to work to drive efficiency, productivity, quality and safety in industrial, manufacturing, infrastructure, public sector and mobility applications.
Hexagon's Manufacturing Intelligence division (Figure 1) enables manufacturers to innovate and create without limits. It drives optimization across the product lifecycle, freeing creators to make better products in innovative ways, for people and the planet.

Figure 1: Hexagon Manufactering Intelligence Logo

Materials Center of Excellence

I was assigned to the Materials Center of Excellence and worked on incubation studies in the field of materials for electric batteries. The goal was to develop and validate certain concepts using techniques such as Machine Learning algorithms to simulate battery performance.

The Materials Center of Excellence (CoE) of Hexagon Manufacturing Intelligence where I was assigned specializes in software and engineering services. It offers innovative materials simulation technology, 100% focused on multi-scale holistic modeling of complex multi-phase materials and structures, as well as materials data management solutions.

The COE's main missions are to help materials suppliers and end users in various industries excel in the following areas:

- Exploit the exceptional properties of heterogeneous materials.

- Improve part performance by mastering the relationships between part manufacturing and material properties.

- Master materials data to support certification processes, collaborative design, CAE (Computer Aided Engineering) simulation and AI/ML.

- Streamline CAE processes by connecting simulation software and associated data.

- Bring innovative and high-quality products to market faster

I work in collaboration with Mr. Moncef Salmi, who holds the role of Business Enablement Leader in the field of AI for materials.

1.2   Scientific context of the internship

1.2.1   Context

Within Hexagon, the strategic framework includes the growing field of electric batteries, which is relatively new and lacks reliable modeling tools. As part of the short-term goals, Hexagon is actively working on providing high-quality modeling solutions in this field. The objectives include collecting existing databases related to electric batteries and exploiting their potential through artificial intelligence-based approaches. The focus is on exploring and demonstrating the significant value that can be derived from a pure AI approach in the field of virtual manufacturing of electric batteries. Finally, the goal is to formulate a comprehensive strategy for the production of these innovative solutions.

1.2.2   Problems to solve

My main responsibilities as stated in my job description at the beginning of your scene were:

1. Conduct a thorough state-of-the-art study of the main AI/ML technologies and algorithms used in the field of electric battery design, examining both academic advances and industry practices.

2. Conduct extensive research on available databases for electric battery design, identifying paid and open-source options, and creating a comprehensive list of these valuable resources.

3. Implement, test, and validate AI models specialized in the automated preprocessing of intelligent data related to the current activity, using machine learning techniques.

4. Design, test, and validate AI models aimed at predicting specific performance of electric batteries, using both open-source AI tools and in-house solutions.

5. Integrate probabilistic AI elements to assess the confidence of predictions.

6. Develop expertise in the recycling/transfer learning method, by efficiently storing and updating a pre-trained model to improve the quality of predictions.

7. Propose innovative ideas for the integration and practical application of the capabilities discovered and implemented within Nexus Materials, with a view to optimizing efficiency and innovation within the company.

1.2.3   Deliverables

The expected deliverables were a PowerPoint presentation and all of my commented codes. In addition to these deliverables, I had to present my research to executives and research teams.

1.2.4   Constraints and available means

The main constraint that arose was the time factor, as the duration of my internship was limited to only 3 months. Since it was my first experience in a company, I had to spend time adapting to the professional environment. Furthermore, the project itself was exploratory, meaning that I did not have access to pre-existing research from Hexagon. I had to start from scratch to develop the project.

2   Scientific Principles Involved

The state of knowledge related to the problem is then described, which leads me in particular to present existing solutions.

2.1   Data Science

The solution I chose to optimize the performance of electric batteries had to be entirely data-driven, that is, it had to be independent of physical modeling. This approach is more promising in terms of results while offering a relatively low barrier to entry. As a result, it can be adopted by a larger number of people. This pure AI-driven method is based on data science, which means that I rely on advanced techniques and algorithms to analyze and best exploit the available data in order to improve the performance of electric batteries.

[1], [4] and [5]

Artificial intelligence
Artificial intelligence (AI) is a branch of computer science that aims to create machines that can replicate some human cognitive functions such as perception, natural language understanding, learning, and problem solving. Machine learning is a subdiscipline of AI that allows machines to learn without being explicitly programmed. Machine learning algorithms can be trained on large amounts of data to detect patterns and establish relationships between inputs and outputs. Deep learning is a branch of machine learning that uses artificial neural networks with multiple layers to perform complex tasks, such as image recognition, language translation, and speech recognition.

Deep Neural Networks (DNNs)
Deep neural networks (DNNs) are computational models inspired by the structure of biological neural networks. They consist of multiple layers of neurons connected together, where each layer performs computations on the input data to produce more complex features.
Traditional neural networks are typically shallow, with one or two hidden layers. DNNs, on the other hand, typically have many hidden layers, sometimes up to several dozen. This additional depth allows DNNs to capture more complex features in the input data. DNNs are typically trained using the **backpropagation** algorithm, which involves adjusting the network weights to minimize the difference between the network's predictions and the target values (the cost function). DNNs typically require considerable training

data to produce accurate results.

Convolutional Neural Network (CNN)

A convolutional neural network (CNN) is a type of neural network that is particularly effective for image recognition. It consists of several layers of neurons, each with the task of detecting features of the image, such as lines, angles, textures, etc.

The **operation** of a CNN is based on the use of *convolutional filters*, which are mathematical patterns applied to the image to detect specific features. Each layer of the network uses a number of filters to analyze the image and extract relevant information.

The CNN also uses *max-pooling* layers, which reduce the size of the image while retaining the most important features. This simplifies the analysis of the image by the following layers of the network and reduces the computation time required.

Finally, the CNN is trained using matrices and their corresponding label (e.g. the fault matrix and the label 'fault-name'). The network then modifies its weights and biases to minimize the prediction error. Once trained, the CNN can be used to predict the label of new matrices.

Vocabulary

Metric: a measure that quantifies the performance, quality, or efficiency of a system, algorithm, or application.

Accuracy: is a performance measure used in machine learning to evaluate the accuracy of a classification model. It is calculated by dividing the number of correct predictions by the total number of predictions.

2.2   Electric batteries

Components and principles

The components and principles underlying the operation of batteries are essential to understand their functionality. First, there are two key electrodes: the anode (negative electrode) and the cathode (positive electrode). Second, the electrolyte serves as a conductive medium for the flow of ions and can exist as a liquid or a gel containing ions. Third, the separator acts as a physical barrier between the electrodes, allowing the transport of ions while preventing short circuits. Finally, current collectors, usually made of metal foils or grids, play a vital role in facilitating the flow of electric current in and out of the battery. Together, these components form the basis of how battery systems work, allowing the storage and release of electrical energy. [**electric-battery**]

Performance measurement

Battery performance can be assessed using several key parameters.

energy density, measured in Wh/kg, quantifies the amount of energy a battery can store. A higher energy density means a greater energy capacity, which is essential for more sustainable energy sources.

lifespan measures the number of charge-discharge cycles a battery can withstand; a longer lifespan indicates that the battery is more durable and reliable.
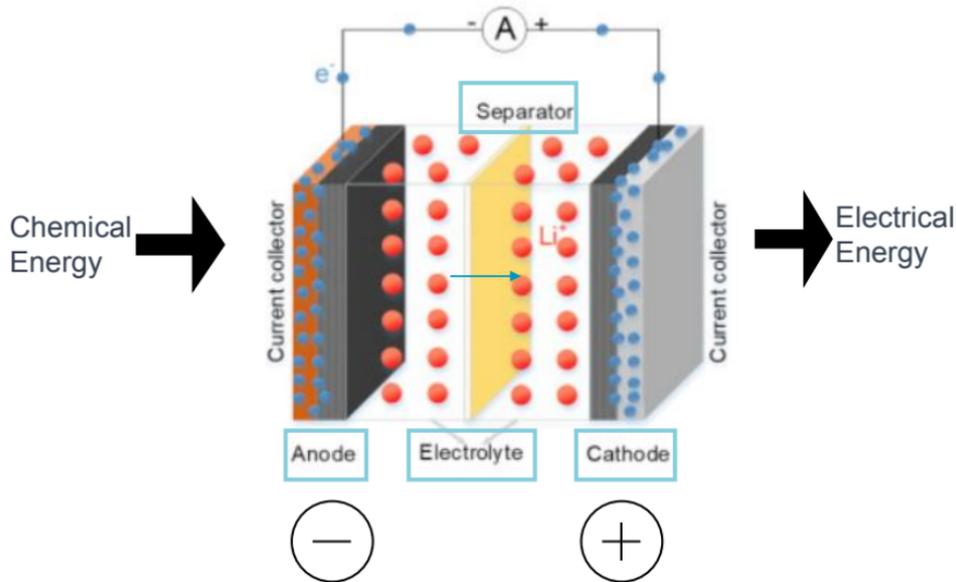
Figure 2: Schematic of an electric battery

efficiency, expressed as a percentage, evaluates how effectively a battery converts input energy into output energy. Higher efficiency means less energy is lost during use, making batteries more environmentally friendly and cost-effective.

These performance parameters play a vital role in evaluating and comparing different battery technologies for various applications.

Key Factors

Several key factors have a significant impact on the performance of electric battery technology:

1. *Temperature* plays a critical role in battery performance. Charging and discharging at an appropriate temperature, typically between 0°C and 45°C for a lithium-ion battery for charging and between -20°C and 55°C for discharging, ensures optimal operation and longevity.

2.*Charge/Discharge Rate (C-Rate)*: The rate at which a battery is charged or discharged, expressed as the C-Rate, affects its performance. It is important to keep the rate at a moderate level to ensure efficient and safe operation.

3. *Internal Resistance:* A lower internal resistance is desirable because it reduces energy loss and heat generation inside the battery, which helps improve efficiency and performance.

4. *State of Charge (SOC):* Monitoring and maintaining the state of charge relative to the battery capacity is essential to optimizing battery life and performance.

5. *Swelling:* Swelling refers to an increase in the size or volume of the battery. It can be

a concern because excessive swelling can lead to mechanical stress, potential damage, and compromised performance.

3   Methods and Solutions

3.1   Methods Implemented

The overall method of work was divided into four main steps:

1. Identify available databases and select the most relevant ones for the project.

2. Perform Exploratory Data Analysis (EDA) on these databases to better understand them.

3. Identify interesting use cases and develop AI algorithms accordingly.

4. Consider integrating the developed algorithms into Hexagon's product Nexus.

Database Search

In the search process, choosing the right Key Performance Indicators (KPIs) is essential to effectively evaluate the data.

1. *Usability:* Usability assesses how usable the database is. In other words, it determines whether the data is modular to be used in training an algorithm.

2. *Value Added:* Value Added evaluates the ability of a database to bring value to Hexagon's products.

3. *Metrics:* The metric measured by a database is essential. A database is interesting if the metric is a factor and/or a performance of electric batteries.

The selection of these KPIs is crucial to determine whether a database is relevant or whether it should be set aside. These measures provide actionable information to make informed decisions and understand the added value of the collected data.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) consists of exploring and understanding the entire data set before applying AI algorithms. This step was necessary and was added to the internship methodology at the request of my tutor. *Python*, with the *Pandas* library, was my choice to perform this EDA because of its ease of use and powerful data manipulation capabilities.

This step allowed:

1. Understanding the data in the dataset I was working with. This involves examining the structure, variables, and characteristics of the dataset.

2. Cleaning the data as inconsistencies, missing values, and outliers are identified and

addressed. Cleaning the data ensures that it is suitable for analysis and modeling.

3. Recognizing patterns, trends, and relationships in the data. I often use data visualization to understand the data by creating various plots and graphs using libraries such as *Matplotlib* and *Seaborn* to visualize distributions, relationships, and trends in the data. The most common plots are histograms, scatter plots, box plots, and bar charts.

AI Algorithms
Data Loading and Preparation
I start by preparing my data: cleaning it, handling missing values, and splitting it into training, validation, and test sets.

Model Design
Next, I sketch out the neural network plan, deciding on the number and type of layers, as well as the activation functions.

Model Compilation
I specify the loss function, optimizer, and metrics to guide the model learning process.

Model Training
Training consists of feeding the model with training data and allowing it to learn from its mistakes over multiple training epochs.

Model evaluation and use
After training, I evaluate its performance on test data to measure the model's performance.

Implementation in Nexus Materials
Nexus Materials is Hexagon's software that allows you to accurately and quickly model new material properties, accelerating the development of next-generation materials. It consists of two main components: Nexus Connect and Nexus Enrich.

- *Nexus Connect* offers the ability to import databases to the cloud, allowing Hexagon customers to access data, whether from their own databases or external databases, using APIs and avoiding manual data transfer steps.

   - *Nexus Enrich* is a preprocessing tool that allows to fill in missing data in materials databases using artificial intelligence techniques.

3.2   Database 1: 3D Battery

Overview
The first database I worked on was called "3D Battery". This database contained X-ray tomography images of battery micro and nanostructures. It included data from 4 different graphite electrodes, which were used in various lithium-ion batteries. These data were published by the Laboratory of Nanoelectronics at ETH Zurich. Each electrode CT scan was associated with a charge and discharge profile. The aim of this database was there-

fore to measure the influence of the nanostructure on the electrochemical performance of batteries.

The database has been linked to a scientific article [2] in which we find the profiles used in Figure 3.
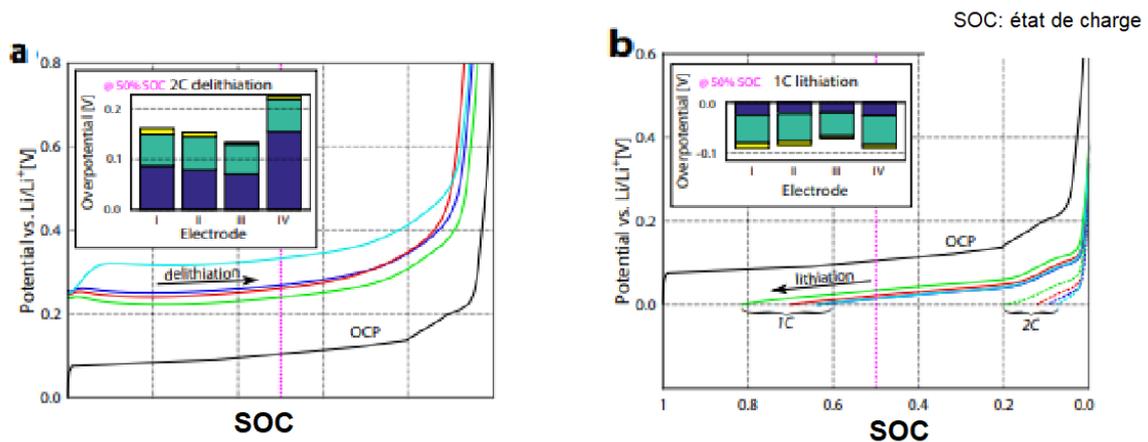

Figure 3: (a) Potential curves for 2C delithiation of a representative 71.5 µm × 71.5 µm × 32.5 Volume µm for the four electrodes (I blue, II red, III green and IV cyan). A state of charge (SOC) of 1 refers to a fully lithiated graphite. The bar plot in the box shows the total overpotential at 50 % SOC for each of the four electrodes. The total overpotential consists of the resistance overpotential (dark blue), the activation overpotential (light green) and the current collector resistance overpotential (yellow). The same is shown in (b) for 1C lithiation. The lithiation curves at 2C are shown as dashed lines. The solid black line marks the open circuit potential of graphite.

Pre-processing

Data preprocessing consists of two essential steps:

- **Data Augmentation**: This first step aims to obtain a database consisting of 60 CT scans from the 4 available electrodes. This data augmentation is crucial to enable efficient learning. The choice of rotation and cropping during this augmentation was made in order to make the data more realistic.

- **Data Extraction**: The second step consisted of manually extracting the data from the profiles present in the scientific article. This was accomplished using the GraphReader tool as shown in Figure 4.

Explanatory Data Analysis (EDA)

During this step, the focus was on understanding the differences between the different curves in order to obtain an estimate of the magnitude of these differences.

Results
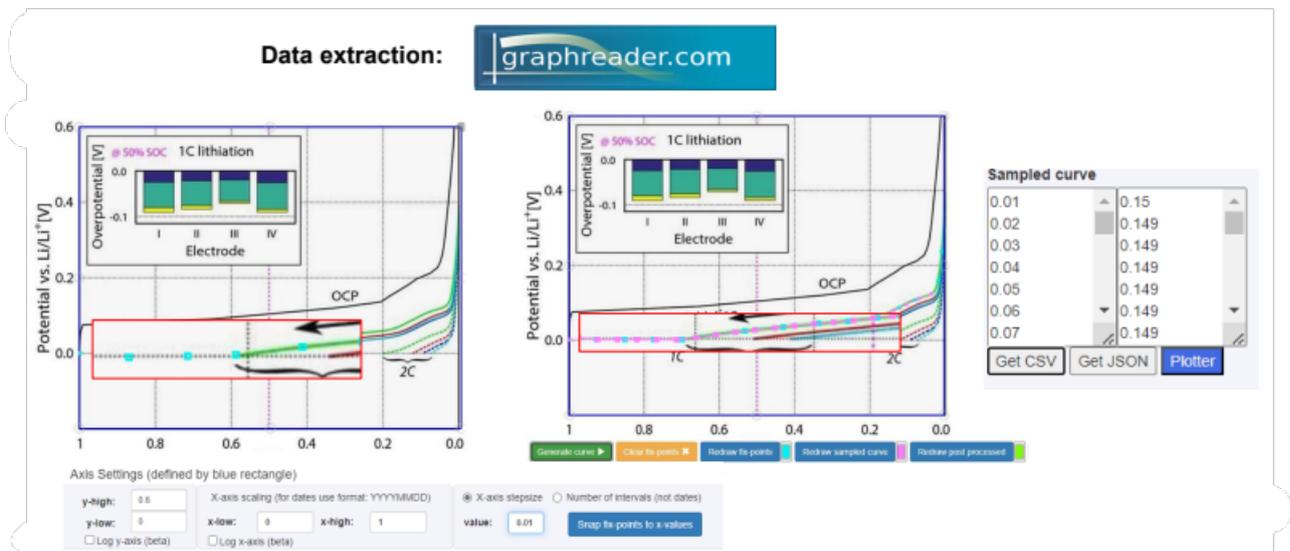
Figure 4: Data extraction method: manual placement of points, parameter adjustment and obtaining the CSV file

- *Open-Source AI Algorithm:* An open-source neural network based on the Keras library was used to perform predictions on the 3D battery images. However, a notable problem was identified: the predictions were identical in all test cases (Figure 5). One possible hypothesis to explain this is that the model may have converged to a local minimum of the loss function. This premature convergence may result in a lack of generalization of the model to the test cases, causing similar predictions. To solve this problem, it may be necessary to adjust the model hyperparameters, modify the neural network architecture, or collect more training data to diversify the learning.
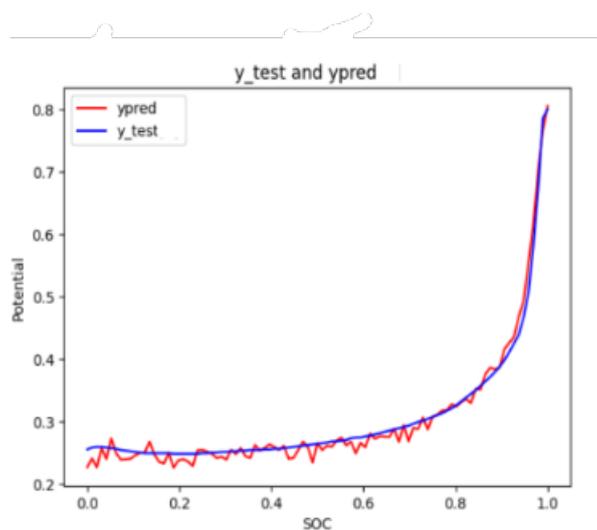
Figure 5: Example of a discharge profile with ypred the predicted curve and y_test the test curve

- *ODYSSEE A-EYE:* ODYSSEE A-Eye is a machine learning solution designed to accelerate product design and development through image- and CAD-based learning. Using images, CAD data, and sensor data as inputs, ODYSSEE A-Eye creates custom AI applications that predict outcomes and data in the field.

  This method proved to be the most promising across all 24 tests conducted (20
  - Average error: 0.0019
  - Minimum error: 0
  - Maximum error: 0.0068

3.3 Database 2: Swelling

Overview

The phenomenon of *Swelling*, particularly observed in cylindrical lithium-ion batteries, is a major issue to consider. It is characterized by the appearance of small spaces between the battery gel coil and its casing. This volume variation can occur after a certain number of charge and discharge cycles of the battery. The Swelling process can have significant consequences on the performance and life of the battery. It is essential to understand this phenomenon and to analyze the batteries before and after a certain number of cycles to assess its impact.

The studied database includes images of batteries before their charge and discharge cycle, as well as images after undergoing different cycles, ranging from 1 to 100 cycles (Figure 6). This approach allows to follow the evolution of the battery swelling over these cycles. This serial analysis of the images offers a detailed view of the impact of the charge and discharge cycles on the battery.
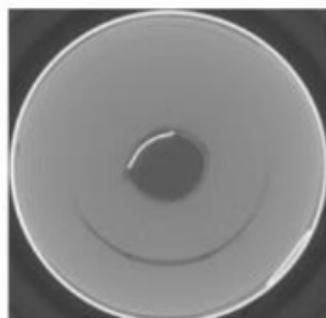
Image de la batterie apès 1 cycle
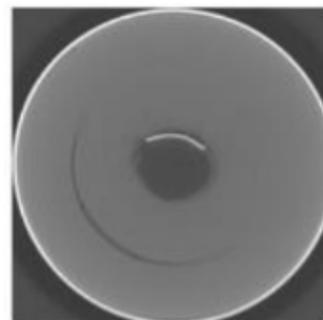
Image de la batterie après 100 cycle

Figure 6: Example of a battery image in the database (Respectively: battery image after 1 cycle and battery image after 100 cycles)

Explanatory Data Analysis (EDA)

The preliminary Explanatory Data Analysis (EDA) revealed that the photos do not have a clear scale, which means that it can be difficult to assess the actual size of the *Swelling* phenomenon. Furthermore, no clear correlation was observed between the number of

pixels in the images and the number of cycles, which makes it difficult to understand the variations related to cycles.

These findings highlight the importance of pre-processing the data before they can be used in a meaningful way.

Pre-processing

- *Manual Labeling*: In this first step, a manual labeling of the images was performed to distinguish the number of cycles in each of them.

- *Image Resizing:* The images were resized to a standard size. This step made the images comparable in size, thus facilitating the comparison of the *Swelling*

- *Volume Measurement:* The volume measurement was undertaken to quantify the *Swelling* of the batteries.
  METHOD 1 - Distance Averaging: This step was performed by averaging the distances in each direction between the resized images. This was done by converting the images to grayscale (Figure 7). The black pixels were then *False* and the white ones *True*. This therefore allows the detection of boundaries.
  The condition to stop is when two consecutive "True" are observed, which may indicate a boundary detection.
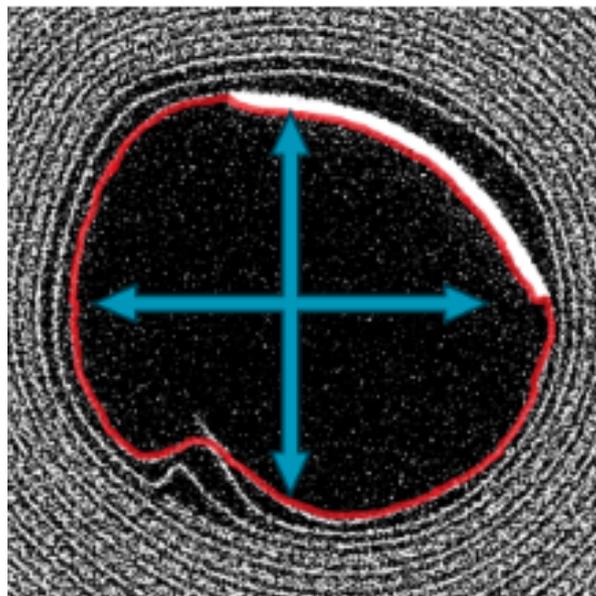
Figure 7: Image converted to grayscale for volume measurement

METHOD 2 - Hough Circle Detection: The Hough Circle Detection method allows to identify circles in an image (Figure 8) , which can be useful to detect boundaries and thus the *Swelling* of batteries. Using Python with the OpenCV (CV2) library, this technique essentially consists of loading the image, reducing the noise, detecting the circles and displaying the results. However, the circle detection cannot be very suitable because it cannot identify some swelling regions in the images due to the deformation of the center of the battery.
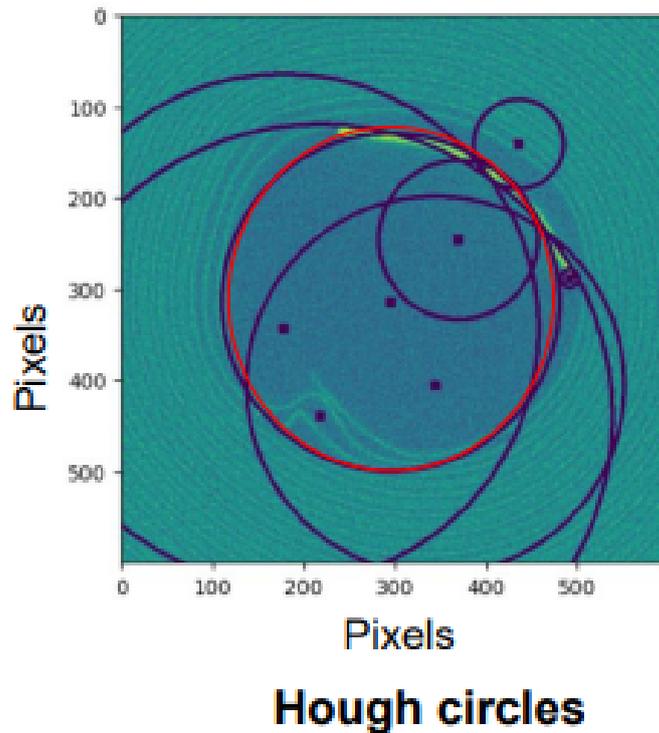
Figure 8: Example of Hough Circle Detection

METHOD 3 - CV2 Contours: The CV2 contour-based method can identify obvious convex shapes in images (Figure 9), which can be particularly useful for spotting *Swelling* areas of batteries. This can give a general idea of the shape variations in batteries over cycles. However, this method can misinterpret the contours and detect shapes that are not related to battery swelling, which can lead to incorrect results.

Pre-processing results
We note that when using the averaging method and CV2 contour detection, we observe a tendency for the volume to decrease inside the boundary at the center of the images. This observation is consistent with the concept of shrinkage caused by the *Swelling* of batteries. In other words, these methods seem to reflect the expected phenomenon, where the *Swelling* of the batteries causes a decrease in the space inside the battery boundary, which is a sign of shrinkage.
However, we note that the Hough circle method, which does not take into account the deformation of this boundary, produces inconsistent results. (Figure 10)

Another example of Swelling: Thermal Swelling

Presentation
In addition to the previous database, I was interested in the influence of the state of charge on thermal Swelling. A database in the form of CSV files, addressed this phe-
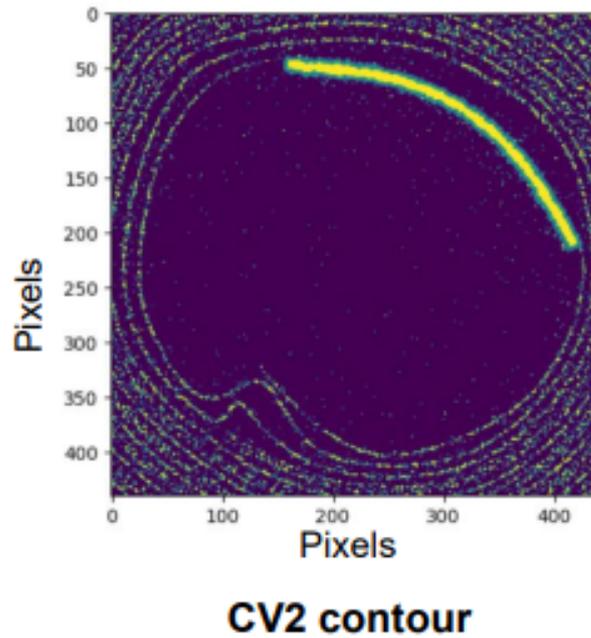
Figure 9: Example of shape detection using the CV2 Contours method
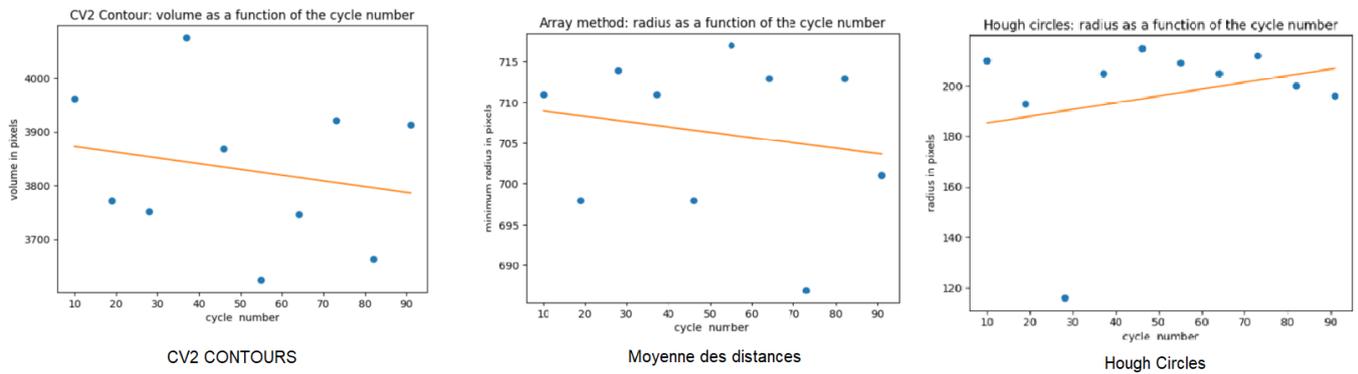
Figure 10: Results of the pre-processing methods

nomenon. The key variables considered were *geographic location*, *temperature* and *state of charge* (SOC) of the batteries. The main objective of this step is to determine how these parameters influence the thermal expansion of batteries. Understanding this relationship is essential to optimize battery life, performance and safety, especially in critical applications such as electric vehicles and energy storage systems. 11

This database has been linked to a scientific article [3].
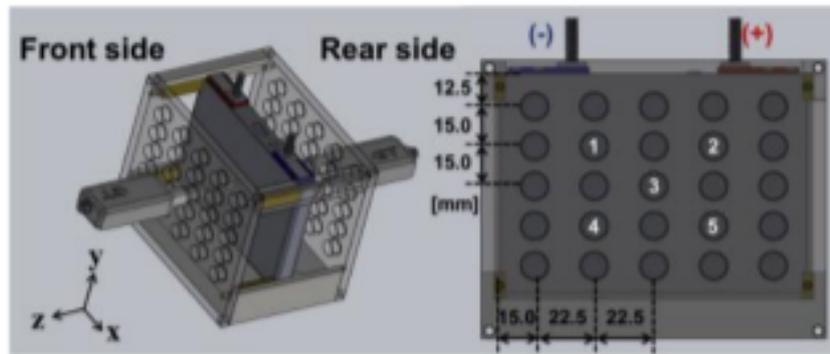
Figure 11: Schematic of the experimental setup showing the device, cell, and sensor locations (1 to 5)
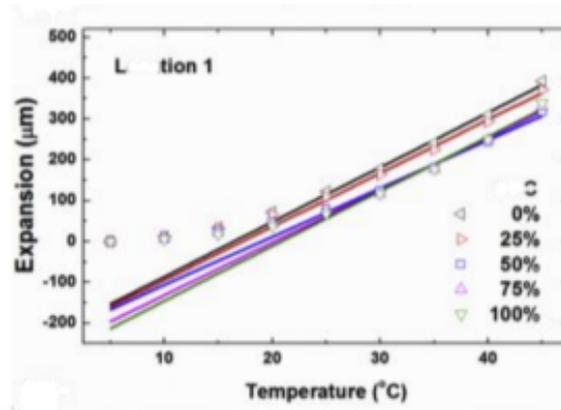
Figure 12: Sample data for location 1

Pre-processing Like the "3D Battery" database, this database required data extraction performed by GraphReader.

Results

A neural network was developed to predict Swelling as a function of temperature and state of charge. The dataset used for this model consists of:
  - Training set: 115 rows, representing 64
  - Validation set: 29 rows, representing 16
  - Test set: 36 rows, representing 20

3.4   Database 3: NASA BATTERY

Overview

This database allows the prediction of the durability of Electronic Batteries and focuses on the evaluation of 34 Li-ion batteries of the 18650 type. The main objective of this study is to measure the state of health (SOH) of these batteries throughout their life cycle until they reach the end-of-life (EOL) criterion which is 70**Pre-processing**

The data presented include variables related to battery characteristics and operations.
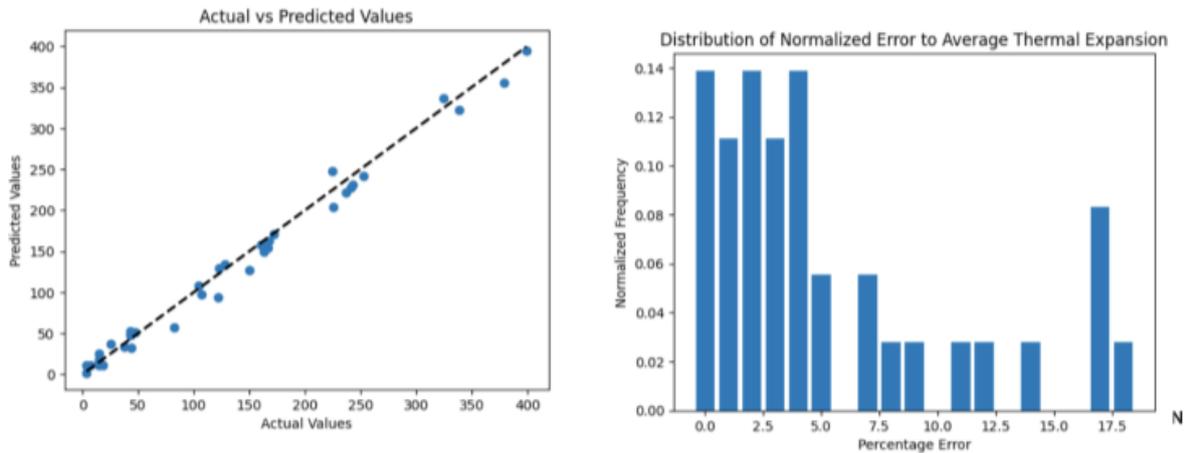
Figure 13: Prediction results
Left: curve of predicted values versus test values
Right: histogram of the distribution of error percentages

Pre-processing was performed to extract from the metadata table, data such as battery name, charge type, charge voltage, discharge type, discharge current, discharge voltage, charge temperature, discharge temperature, capacity, state of health, and number of charge and discharge cycles.

In addition, the values were in text files. Therefore, another step was necessary to automatically extract this information.14

Figure 14: Extracting data from the ReadMe file

Understanding the importance of the end-of-life criterion at 70

Pre-processing Results

In order to arrive at , the strategy was to adopt a complementary approach by using an open-source neural network (Keras) with Hexagon's ODYSSEE CAE Lunar tool. My main goal was to accurately predict the number of cycles a battery can perform before reaching the defined end-of-life (EOL) criteria. This complementary approach therefore allows me to determine the health profile as a function of cycles by knowing the shutdown
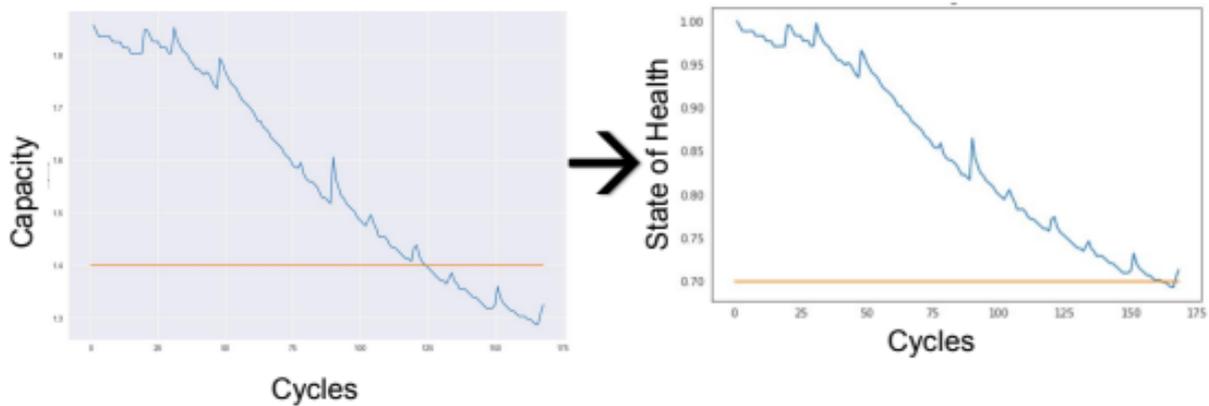
Figure 15: Example of Converting Capacity to State of Health

limit. The algorithm was not 100% developed.

Nevertheless, by starting this approach, pre-processing results were observed.

To address the problem of profiles with physically impossible responses that should decrease steadily but exhibit fluctuations, an additional pre-processing step was integrated. This step consists of smoothing the curve to eliminate erratic fluctuations while preserving essential information. In addition, outliers are handled by replacing them with the moving average value.
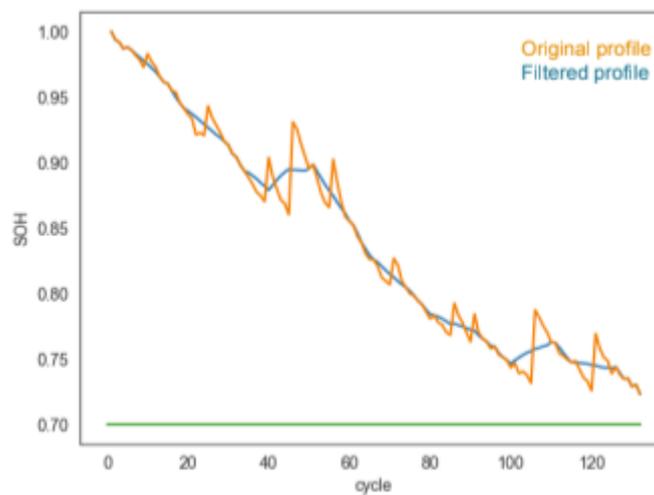

Figure 16: Example of smoothing physically impossible curves

3.5   Implementation in Nexus

The overall idea is that each of these algorithms can be part of Nexus Materials. They can be integrated as algorithms within a new product 'Nexus Design', which is itself associated with Nexus Connect and Nexus Enrich. This concept allows the customer to use the algorithms pre-trained on open source data to predict the performance of battery

models in real time, via the cloud. In addition, the customer also has the possibility to import their own datasets, fill in the information gaps and use them to re-train the algorithms. This amounts to developing a concept of transfer learning, where algorithms can capitalize on pre-existing expertise and adapt to the customer's data for personalized and accurate predictions. This approach offers considerable flexibility and customization to meet the customer's specific needs in the field of battery sustainability.
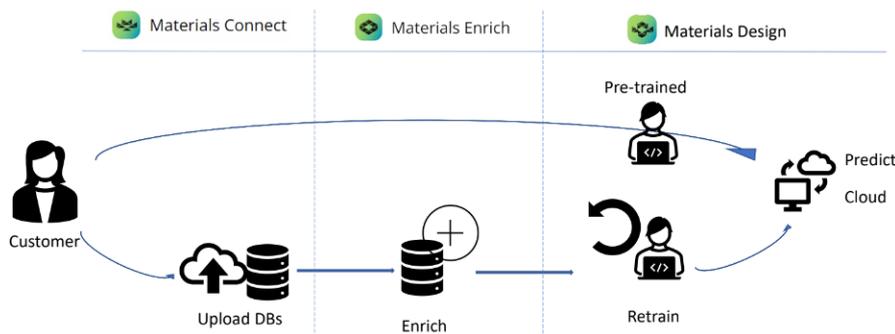

Figure 17: Using Nexus Design

4   Conclusion

Domain	Result	Learning	Market Opportunity for Hexagon
Databases on electric batteries	There are various resources, but useful data is scarce	We can rely on open-source databases to show the potential of AI modeling, but we don't have enough data to train and promote mature models	Battery databases can be a value proposition in itself, in addition to providing value to customers who have already reached a sufficient number of BD
Data preprocessing	The available data is often unstructured and unusable in its raw state.	A large part of the efforts is devoted to data processing: Word processing, image processing, etc.	An efficient data preprocessing tool for battery data electronics can be a very interesting and differentiated value proposition tool
AI modeling	Odyssee A-Eye has achieved good results on computed tomography (CT-Scan) data	Odyssee A-Eye presents an interesting differentiated value proposition when little image data is available	Focus on Odyssee A-Eye as a strategic AI modeling tool for CT data CT data for electronic batteries

References

[1] Gao Huang et al. "Deep Networks with Stochastic Depth". In: *Publisher: arXiv Version Number: 3* (2016). DOI: `10.48550/ARXIV.1603.09382`. URL: `https://arxiv.org/abs/1603.09382` (visited on 01/26/2023).

[2] Simon Müller. "Quantifying Inhomogeneity of Lithium Ion Battery Electrodes and Its Influence on Electrochemical Performance". In: *J. Electrochem. Soc. 165 A339* (2018).

[3] Ki-Yong Oh and Bogdan Epureanu. "A novel thermal swelling model for a rechargeable lithium-ion battery cell". In: *Journal of Power Sources* 303 (Jan. 2016), pp. 86–96. DOI: `10.1016/j.jpowsour.2015.10.085`.

[4] Josh Patterson and Adam Gibson. *Deep learning en action: la référence du praticien.* fre. Paris: First interactive O'Reilly, 2018. ISBN: 978-2-412-03744-7.

[5] Bharath Ramsundar and Reza Bosagh Zadeh. *TensorFlow pour le deep learning: de la régression linéaire à l'apprentissage par renforcement.* fre. Paris: First interactive O'Reilly, 2018. ISBN: 978-2-412-04116-1.