Engineering Applications of Artificial Intelligence 114 (2022) 105120

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

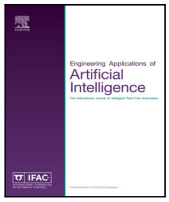

Image- and health indicator-based transfer learning hybridization for battery RUL prediction

Jonathan Couture, Xianke Lin *

Department of Automotive and Mechatronics Engineering, Ontario Tech University, Oshawa, Ontario L1G 0C5, Canada

ARTICLE INFO

Keywords:
Transfer learning
Neural network hybridization
Lithium-ion battery
Health prognostics prediction
Remaining-useful-life

ABSTRACT

With the recent influx of electric vehicles and other electrical consumer products, the market has created a large demand for lithium-ion batteries, however, the non-linearity of their internal aging mechanisms still makes them difficult to predict. This study proposes a novel hybridized machine learning model that aims to be accessible to all researchers working in the field of battery prognostics prediction. To make it accessible to all, this study prides itself on using pre-trained and publicly available neural networks. To make full use of these available transferable networks, the traditional battery prognostics prediction methodology of using health indicators is hybridized with the use of images of the raw data curves. Also, this study tests the accuracy performance of the models by using images by themselves without the traditional health indicators to provide a proof of concept that these can be used independently as well An improvement of 6.72% in the remaining-useful-life (RUL) prediction accuracy was observed by simply adding image-based inputs, without the need for additional extensive pre-processing to an already existing health indicator-based model for the prediction of batteries' RUL, highlighting the benefits of image-based inputs for use in regression tasks.

1. Introduction

With more and more governments enticing the use of electric transportation, we can expect a large influx of electric vehicles, powered by lithium-ion batteries, to enter the market in the coming years. Nonetheless, the methodology of predicting the remaining-useful-life and state-of-health of these batteries has had a hard time catching up with the rapid advancements in battery technology. Battery management systems play a more important role than ever in our everyday lives, and one of their biggest challenges is still to determine important health factors from the partial information available to them.

This study seeks to propose a novel approach that could easily be implemented alongside existing battery prognostic prediction methodology with little added computational and physical burden. Using existing pre-trained neural networks that are publicly available to all through the popular open-sourced PyTorch machine learning library, this paper prioritizes the ease of implementation of its method. To allow the use of pre-trained networks, this study makes use of the already standard method of transfer-learning, which was originally proposed back in 1976 (Bozinocski and Fulgosi, 1976) but which has never seen much use in regression applications and has only been picked up more and more in the recent years.

Transfer learning is a method of problem-solving that mimics the humanlike learning strategy where we use information gained from previously solved tasks and apply this learned knowledge to related applications (Kocer and Arslan, 2010). As mentioned, these approaches have had limited usage in regression applications since most publicly available pre-trained neural networks are for classification applications, such as data labeling and object detection. These classification tasks use images as their inputs and classes as their outputs, whilst the majority of regression tasks use numerical data to obtain a single numerical output. This requires that the fully-connected and output layers of the transferred networks be retrained to make use of the models' feature extraction abilities.

Using images as inputs provides the benefit of showing a large amount of data in a concise and computationally efficient manner while providing the data filtering abilities of convolutional layers. More traditionally used networks for this task, such as the long short-term memory networks (LSTMs) and fully-connected neural networks (FCNNs), have poor input filtering, making them prone to large errors if faulty information is sent through the network. Convolutional neural networks (CNNs) can learn which points on an image are more prone to noise and error, and which are more reliable within the obtained data. This study will present various implementations which can improve the prediction accuracy of existing methodology with the use of limited additional computational requirements. To do so, we propose various transfer learning models to be used by themselves or with their hybridization alongside traditional health indicators. The method is then tested against a benchmark value, obtained by training an FCNN

* Corresponding author.
E-mail address: xiankelin@ieee.org (X. Lin).

https://doi.org/10.1016/j.engappai.2022.105120
Received 13 April 2022; Received in revised form 12 June 2022; Accepted 21 June 2022
Available online 1 July 2022

Nomenclature

Abbreviations

CNN	Convolutional neural network
DCNN	Dilated convolutional neural network
FCNN	Fully-connected neural network
HI	Health indicator
LFP	Lithium iron phosphate
LSTM	Long short-term memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MES	Multiple expert systems
RMSE	Root mean squared error
RUL	Remaining-useful-life
TLPH	Transfer learning parallel hybrid

Variables

$dQdV$	The capacity of the battery in terms of the voltage
$volt_t$	The time that is taken to reach 3.3 V from 3.15 V during charging
$curr$	Average current of the final shown cycle
Q_c	Charging capacity
Q_d	Discharging capacity
IR	Internal resistance
T	Temperature
ρ	Pearson's correlation coefficient
σ	Standard deviation

Suffix

$mean$	The average value of the observed sample
min	The minimum value within the observed sample
max	The maximum value within the observed sample
var	The variance of the observed sample
lin	Linearly interpolated
d	Deviation of the variable over the shown cycles of data

with health indicators. The networks are then tested using 5 and 10 consecutive cycles of data, and due to the more intuitive learning of image-based networks, a single cycle of data is tested as well.

Making use of these pre-trained models is greatly beneficial when the available datasets are small since the number of the network's trainable parameters is significantly reduced. Also, by reducing the number of trainable parameters, there are significant time savings to be had for offline operations, where computational hardware might be limited and continuous machine learning is sought. Additionally, important time can also be saved from the pre-processing labor usually required to develop complex neural network architecture, as well as for the expert knowledge required to extract useful health indicators from the available data. By using already available machine learning architectures and allowing these to filter the raw data, the usefulness of the proposed methodology has a very wide range of applications.

The remainder of this paper will be structured as follows: Section 2 will enumerate the contributions of this study and Section 3 will discuss related literature and existing methodologies. Section 4 will then describe the proposed method, followed by a correlation analysis in Section 5 to determine the inputs to the neural networks. The results will be presented in Section 6, then a conclusion in Section 7, and then a short discussion regarding future work in Section 8.

2. Contributions

The main contributions of this study are as follows:

(1) Introduction of image-based inputs for regression applications. These images are created with the use of existing data and provide the benefit of showing the entire battery's charging and discharging cycles within a single input. This approach has the advantage of requiring very little preprocessing and little to no required knowledge in the field;
(2) A comparison of various transfer learning models is made for use in battery remaining-useful-life prediction. A set of optimal inputs are obtained and transformed into images to be used with the transferred models. With this, the networks can predict the RUL of a battery when shown only the data of a single charging and discharging cycle;
(3) A novel architecture is proposed, consisting of a transferred image-based neural network, alongside numerical health indicators. The proposed method called the transfer learning parallel hybrid (TLPH) method can improve existing methodology by using the combination of image-based inputs with pre-existing health indicators, allowing this method to be easily implementable, and completely open-source. Assuming that voltage and current values were already available for the generation of health indicators, no additional data is necessary to implement this method.

3. Related works

The majority of the work published on the prediction of battery health prognostics depends heavily on the quality of the health indicators (HIs) used as inputs, with papers solely addressing the different health indicators and their impact on the prediction accuracy (Liu and Xu, 2019). The most popular HIs are based on the voltage, temperature, and current, or a combination of these. Most new research on RUL prediction typically involves the use of a new indicator or a different neural network architecture. The main benefit of image-based machine learning is its ability to look at entire data sequences and letting the network learn which parts of it are more relevant to the desired output, effectively automating the HI extraction process. The sections that do not contain as much useful information are simply ignored by the network without forcing the data scientist to manually remove them from the inputs.

Oftentimes in the literature, NASA's battery degradation datasets (Saha and Goebel, 2007) are used (Liu et al., 2015; Ren et al., 2021; Zhou and Huang, 2016). However, these datasets are not utilized within the scope of this study due to their limited number of batteries, having only the lifecycle data of 62 batteries. Since machine learning algorithms are very data-hungry by nature, using such a low amount of data could easily lead to overfitting. Within the literature, some studies were performed using only 2 distinct cells, which can clearly become problematic and from which very few conclusions can reasonably be made in regards to the overall performance of the methods proposed. Because of this, the MIT/Stanford lithium iron phosphate (LFP) battery dataset (Severson et al., 2019) is used, which includes lifecycle data from 124 LFP cells, which should allow our method to become significantly more robust and reliable.

A very popular neural network within the literature for RUL estimation is the long-short-term memory (LSTM) network, which was used a great deal (Li et al., 2019; Liu et al., 2021; Zhang et al., 2018) due to its ability to learn order dependence between items in a sequence, making it ideal for time-dependent sequential data. However, its main drawback is its lack of input filtering, making it vulnerable to outliers within the inputs which could significantly throw off the

model or would require the data scientist to perform a large amount of preprocessing before training. Due to this drawback of the LSTMs, there is an abundant amount of research published regarding the optimal battery health indicators to use as inputs, which would allow the network to see the most useful information reliably. Also, from this, various novel implementations have suggested the hybridization of LSTMs with convolutional neural networks (CNNs) to make use of their input filtering abilities. These networks, being the backbone of modern image and speech recognition and language translation algorithms, can look at much larger datasets and adjust their filter weights to learn the useful patterns that are present within the data. By learning where the useful information lies, CNNs are also able to learn to disregard the unvaluable sections of the inputs, which constitutes their filtering abilities. Hybridizing these networks prior to an LSTM allows the overall method to consider time-dependent sequential data with the benefits of input filtering (Ma et al., 2019; Ren et al., 2021; Zraibi et al., 2021). From these implementations of CNN hybridization came the idea to feed images of data curves for regression applications.

3.1. Convolutional layers

As of the writing of this paper, all publicly available networks on the PyTorch platform are image-recognition-based models. Traditionally, to develop an image-based model, you first need to make use of convolutional layers, which act as filters that go over the input image and extract useful information. Layers closer to the image extract low-hierarchical features, such as lines, corners, and curves, whilst the deeper the network goes, the more complex the extracted features can be, such as a car or a house. To make use of these, images are created for this study to be used as inputs to these transferred networks.

These convolutional layers work by applying filters, or kernels, onto a given portion of the image, where it extracts the relevant information, and then moves on to a different section of the image. This new filtered information becomes the feature map, which is then filtered by the following convolutional layer. Every time filtering occurs, the feature map reduces since the filtering process applies the sum of an element-wise multiplication between the kernel and the pixel values of the input. Traditionally, the feature map ends up being used as inputs to the fully-connected layers, which then learn to map the different features seen in the image to their respectable object class. In the case of this study, these fully-connected layers instead learn to transform these features into a single RUL prediction.

3.2. Transfer learning

Traditional machine learning techniques seek to learn each task from scratch, which requires a large amount of data and makes them very prone to overfit to the training dataset. Transfer learning instead uses previously obtained knowledge, acquired when training on a previous task, and adapts it to a new target task, requiring significantly less training data. Depending on whether the transferred network is re-trained or not, the overall methodology can also become considerably less computationally expensive.

This method works by transferring over the weights and biases of an already trained neural network to initialize a new model that uses the same architecture (Kocer and Arslan, 2010). In the scope of this study, we imported image recognition neural networks to treat our image inputs, since these models have already been trained on large image datasets and have already learned to recognize basic features. Also, since our output is different from the trained models, the final fully-connected layers are replaced with new ones, which in turn reinitializes their weights and biases, requiring them to be trained on the studied dataset. This allows the first convolutions to extract features from the input image, and the final layers to triage these features and transform them into RUL predictions. Other works in research fields other than battery prognostic predictions, such as Roy (2022), managed to successfully make use of transfer learning techniques for classification applications.

3.3. Research gaps

From surveying the available battery prognostic prediction literature, it can be observed that the vast majority of proposed methodologies require a large amount of time-intensive pre-processing labor and expert knowledge within the field. These requirements have not been properly addressed to this date, causing a barrier for less experienced researchers and users willing to implement some of the proposed methodologies. There is also still a great need for a large amount of accurate and reliable datasets, where a real-life electric vehicle dataset should be prioritized. Such a dataset would help standardize future methodologies by allowing researchers to use the same data and thus allow a fair and accurate comparison between the studies. The computational complexity of the proposed methods within the literature is oftentimes a neglected factor but plays an important role within real-life applications due to the potential for limited hardware computational complexity, especially for offline applications. Also, there is a lack of standardization when comparing the accuracy metrics of the various presented models, which in turn hinders a fair comparison between proposed methodologies.

Additionally, limiting the battery prognostics field to only numerical inputs seems to have bottlenecked many recent studies into proposing new or hybrid health indicators to try and improve the performance of the proposed models. Accessing the work provided by the wide range of experts working on image classification tasks has a huge potential for the future of battery prognostics predictions as well as other regression fields.

4. Proposed method

This paper firstly uses a simple artificial neural network architecture to act as a baseline for the proposed methodology and also to validate the health indicators that are used. Transferred networks are then trained on an image dataset to prove their ability to perform well in regression applications. The health indicators are then combined alongside the outputs of a transferred network to obtain a transfer learning parallel-hybrid architecture (TLPH). Traditional methods might call for the use of an LSTM neural network due to its strong prediction abilities of time-dependent variables, but due to the minimalistic nature of the proposed method, this study opted to discard this architecture in favor of incorporating time-sensitive information within the health indicators themselves. These inputs can then be directly combined with the transferred network's output without the use of LSTM layers within the final TLPH network.

4.1. The dataset

The MIT/Stanford battery dataset (Severson et al., 2019) was used for the training and testing of this study's machine learning implementation. This large dataset consists of 124 LFP cells that were cycled to their end of life under various operating conditions. The end-of-life of the batteries is defined as a 20% capacity fade from the nominal capacity, which is a standard value used within the literature (Casal et al., 2019). The useful life of the cells varied between 148 and 2237, however, we have decided to only include the cells that had cycle lives between 400 and 1175 to reduce the variance within the data. A total of 14 cells were excluded as a result of this process.

Again, for the sake of variance, we randomly excluded cells within the 400–500 cycle life range. If left as is, the neural network might favor low cycle life batteries by performing better on them and slightly worse on the rest. It would be easy to leave the dataset as it is and test on more of these batteries, but although the network's accuracy would look good, the overall model would not be as robust. For this reason, we removed another 16 cells from the dataset to balance the variance. Fig. 1 shows the remaining number of batteries that were used in this study.

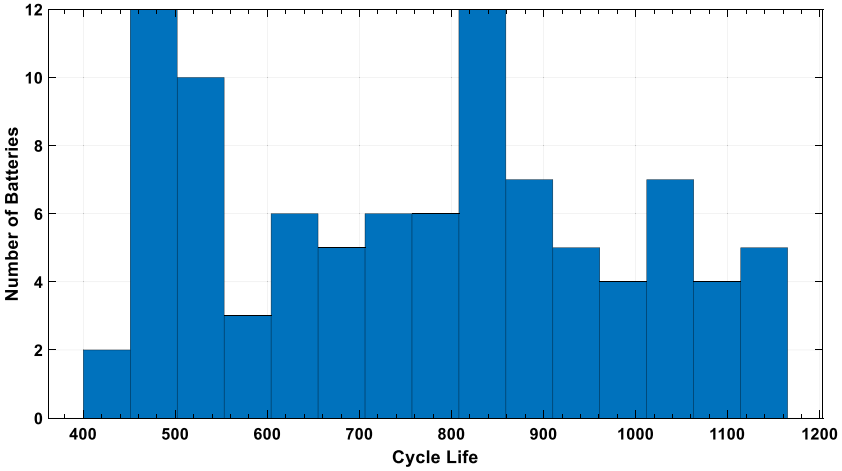

Fig. 1. Cycle life variance for the studied dataset.

For the testing set, 20% of the remaining batteries were chosen randomly to guarantee the robustness of the network. A total of 19 randomly chosen cells are used for the testing of the proposed methods, and 75 cells are used for the training of the networks.

4.2. Accuracy metrics

This section highlights the different accuracy metrics used to assess the performance of the proposed methodology. To facilitate the comparison with other methods, we used a few different metrics. For this study, the mean absolute percentage error (MAPE) is gathered, alongside the mean absolute error (MAE) and the root mean squared error (RMSE). These are pretty standard accuracy metrics used to assess the performance of neural networks because they represent the overall prediction accuracy, alongside the average error, in cycles, and the standard deviation of the residuals, also in cycles, respectively. Showing only one or two of these metrics only presents the reader with part of the method's performance, especially since the MAPE can easily be manipulated by showing the network only high target values and omitting to predict values when the true value is low.

$$MAPE(\%) = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right| \tag{1}$$

where A is the actual target value, F is the predicted value, and n is the number of predictions made. When looking at the equation, it should be noted that low target values are more penalizing to this metric than high target values, making it imperative that all results obtained are considered to provide an accurate MAPE value.

The MAE, given in cycles, represents the mean prediction error. However, being a useful metric on its own, the average prediction error is still just an average value, meaning that it does not consider the fact that maybe the network is only really good at predicting small values, but considerably worst at predicting large values, or vice-versa. This scenario would ultimately achieve a reasonable MAE but neglects to examine the robustness of the method under different prediction ranges. This is why the RMSE is also used in this study.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|A_i - F_i| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - F_i)^2} \tag{3}$$

where A is the actual target value, F is the predicted value, and n is the number of predictions made. This final accuracy metric is very useful because it penalizes more harshly large errors by squaring its value and is ultimately a good indicator of the robustness of the network under different target ranges.

5. Correlation analysis

For this study, the Pearson's correlation coefficient is used to determine the optimal inputs to the neural network. This coefficient serves to compare the linear correlation between two sets of data by using the following equation:

$$\rho(A,B) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{A_i - \mu_A}{\sigma_A}\right)\left(\frac{B_i - \mu_B}{\sigma_B}\right) \tag{4}$$

where μ and σ are the mean and standard deviation, respectively, of the variables A and B. The maximal values possible for the correlation coefficient are -1 and 1, representing a perfect negative and positive correlation between the two studied variables. A value of 0 would indicate that there is absolutely no linear correlation between the two studied variables.

5.1. Correlation results

A few variables were examined within the scope of this study to obtain an optimal set of health indicators. For this, we compiled the correlation coefficients between the given variables and the RUL for each cell in the dataset and then obtained the average correlation value. To make sure that the correlation would be consistent throughout the studied cells, we also obtained the standard deviation of the correlation results to determine which variables were consistently close to their median correlation value. Even with a high average correlation, the results could be thrown off if the testing set diverges too much from the training data.

A few of these examined health indicators were proposed in the original paper presenting this dataset (Severson et al., 2019), such as $dQdV$, Qd_{lin}, and Td_{lin}. Summary statistics are then calculated for most indicators, such as the mean, minimum, and variance values to determine their correlation with the RUL. All of these statistic values are scalar values capturing their change between 5 cycles. The $volt_t$ health indicator was obtained from Hong et al. (2020) and represents the time taken to reach 3.3 V from 3.15 V. The following figure shows the obtained correlation coefficient results, averaged over all of the cells, for each health indicator when compared with the RUL of the cells (see Fig. 2).

A threshold value of 0.6 was chosen to separate the health indicators into low- and high-correlation groups. It is clear that there are a few indicators that managed to achieve satisfactory correlation value, but since these were averaged over all of the batteries in this study, it is imperative to also take a look at their consistency from cell to cell. For this, the standard deviation is used to determine whether the

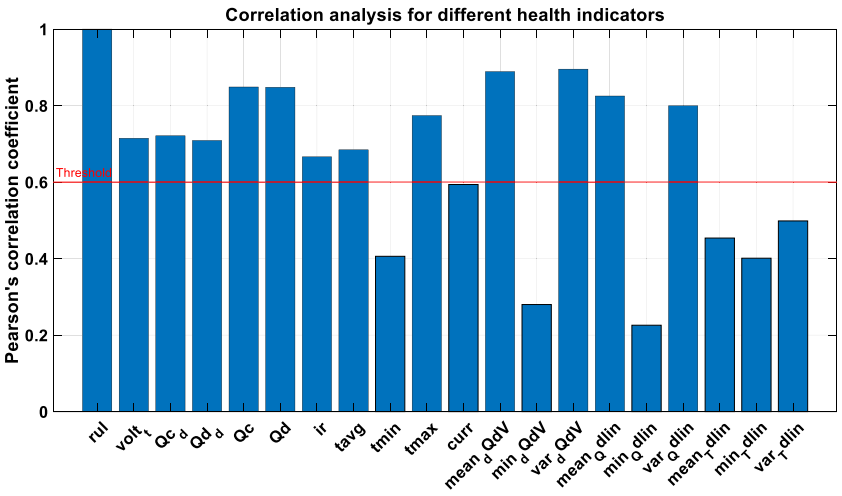

Fig. 2. Pearson's correlation coefficient for the compared health indicators.

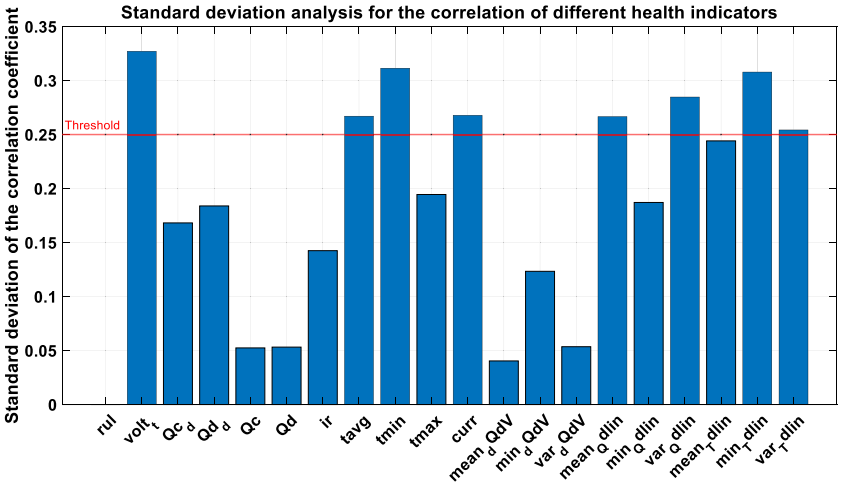

Fig. 3. Standard deviation results of the correlation coefficient throughout the studied dataset.

correlation values are consistent. The following equation was used to obtain the standard deviation between the variables.

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} \tag{5}$$

where σ is the standard deviation within the observed variable, x_i is the ith value of the x-sample, μ is the mean value of the x-variable, and N is the size of the population. The following figure presents the deviation results obtained (see Fig. 3).

For our purpose, a low standard deviation is sought after, since it indicates that the correlation coefficient is consistent throughout the studied battery cells. A threshold value of 0.25 is selected to again separate the unvaluable indicators. The chosen health indicators that are used throughout this study are presented in Section 5.2.1.

5.2. Neural network inputs

5.2.1. Health indicator inputs

A total of 12 health indicators are chosen and used as inputs to the fully-connected neural network for training. These inputs are not time-dependent, meaning that they do not need to be sequentially fed to the network, and so an LSTM network is not necessary for this task. The chosen health indicators for this study are:

- The average current over the final cycle of data shown, *curr*, in A, which will ideally translate to the charging protocol used on the battery, since there is a large variety of these protocols within the dataset which all have their effect on the batteries' degradation;
- Qc_d and Qd_d, in Ah, represent the change in charging and discharging capacity, respectfully, over the shown cycles of data;
- Qd is the discharging capacity value, in Ah, of the final cycle show. The charging capacity has been discarded due to its high similarities with Qd which would have been redundant information;
- IR, which is the internal resistance value at the final cycle shown;
- T_{max} and T_{avg} are, respectfully, the maximum and average temperature values over the final shown cycle. We opted to include both of these metrics since there was a significant amount of outliers within the temperature values which could easily throw off the network with a single high maximal temperature value, ideally, the network will learn to disregard these outliers with the help of the average temperature during the cycle;
- As mentioned previously, $dQdV$ and Qd_{lin}, which were first presented in Severson et al. (2019) reflect the capacity of the battery in terms of the voltage and the linearly interpolated discharging capacity, respectively. We opted to include the mean and standard deviation values of both of these health indicators due to their high correlation with the RUL;
- We also opted to include the $volt_t$ even though its correlation coefficient had a high standard deviation due to its success in the previously mentioned study (Hong et al., 2020).

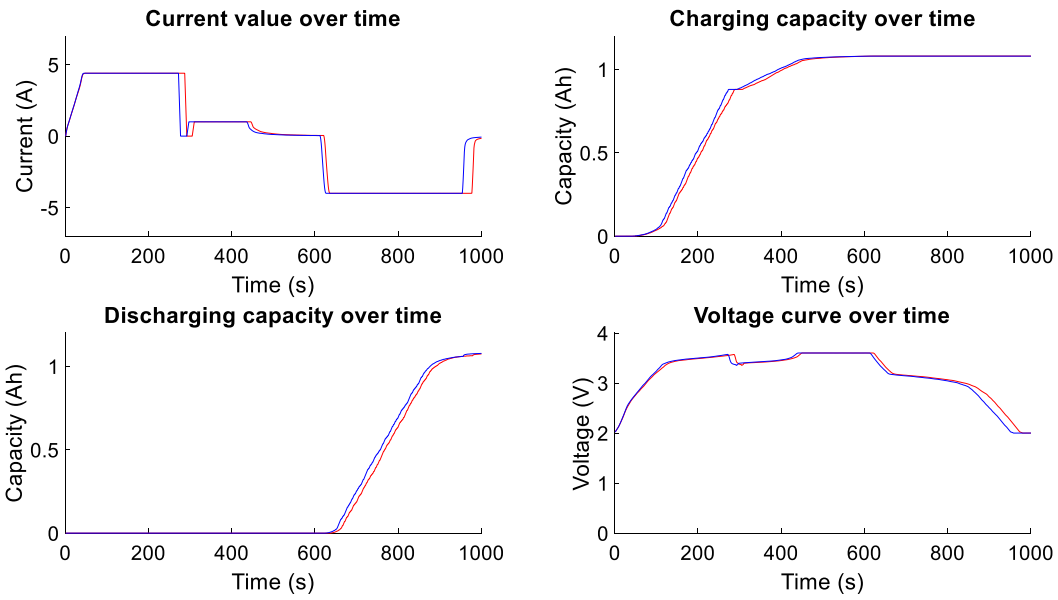

Fig. 4. Visual representation of the degradation occurring within the image's input curves.

For the testing of the 1 cycle dataset, only 10 health indicators are used as inputs, since both Qd_d and Qc_d are removed due to having no known degradation between consecutive cycles.

5.2.2. Image inputs

Since this study seeks to allow the easy implementation of its methods, we opted to use data that should not require much preprocessing labor, unlike the previously mentioned health indicators, which need to be manually extracted from the data. For this, the current, capacity, and voltage data are transformed into images. These represent, respectively, the flow of the electric charge to and from the battery during charging and discharging, the total amount of energy being stored inside the battery, and the difference in potential between both electrodes.

Here, the current is used as an input since it is a readily available measurement gathered by the onboard battery management system. For this study, both the charging and discharging data were used to more accurately compare our methodology with the available literature. However, the charging data would typically be preferred for an electric vehicle scenario due to the discharging conditions being highly volatile depending on the user's driving habits. Since the studied number of cycles for this paper is 1, 5, and 10 consecutive cycles of data, we opted to only show the network the first and last cycle to prevent any significant overlap between the curves. Of course, for the 1 cycle dataset, only a single curve is present in the image. The following figure shows an example of the degradation information present in the images of 5 consecutive cycles (see Fig. 4).

Since the neural networks would care little for the titles and axis value, these are removed to allow for more room to expand the curves themselves. To allow the network to learn the baseline of a healthy and degraded cell, the axes are fixed between set values so that these do not change to center the curve in the image, instead, the curves will translate on the axis, allowing the neural network to use their position as useful information. The curves in blues represent the data from the n_{th} charging and discharging cycle, and the curves in red show the data from the $n_{th}+5$ studied cycle. The 10 cycle dataset will use the $n_{th}+10$ cycle and the single-cycle dataset will only use the n_{th} cycle. Fig. 5 shows an example of the workflow that is used throughout this study.

Each generated image was saved as a 256 by 256 pixels image to allow the use of the various transfer learning networks. Due to the nature of convolutional layers, the input images need to be of a minimum size, since the filter matrixes reduce the feature map at every iteration, with the exception of 1×1 filters, which are not commonly used. It would be possible to increase the pictures' size by allowing more pixels to be present, but this would come with the cost of an exponentially higher computational burden to load and process these images.

6. Results

To prove the accuracy and efficiency of our proposed methodology, three studies were extracted from the literature for comparison. Mainly, the Elastic Net, proposed by Severson et al. (2019), made use of regularization techniques to simultaneously perform model selection and fitting. For hyperparameter optimization, four-fold cross-validation and Monte Carlo sampling are applied. The main difference between this methodology and the one presented within this paper is the inputs to the networks. Mainly, Severson et al. proposed to make use of the first 100 cycles of the battery's lifetime instead of randomly selected cycles throughout the battery's life. This in turn causes the network to have significantly less variability to learn within the data and also gives it a very large amount of inputs, both factoring in the high accuracy obtained. Within real-life electric vehicle scenarios, it is unlikely that such historical data would be available to the battery management system, which is why this present study opted to select random cycles throughout the batteries' life cycle. Secondly, the dilated convolutional neural network (DCNN) proposed by Hong et al. (2020) uses this architecture to filter raw cycling data to extract features through the CNN which are then fed into fully connected layers to obtain an RUL prediction. Their proposed architecture is quite similar to the one from this study, however, the difference comes from the inputs themselves. Where our proposed methodology uses images of curves in the form of pixel arrays, Hong et al. used the numerical arrays for their various inputs. Finally, the model proposed by Sanz-Gorrachategui et al. (2021) will also be used for comparison. These authors used several health indicators to first classify the observed cell into a long- and a short-RUL and then used the appropriate 'expert' model for prediction. The idea behind this model was that this course of action would limit the overall variability that each of the regression networks would have to learn, which in turn would provide an increased prediction accuracy. This combination of various smaller models is combined into a single prediction pipeline and referred to as the Multiple Expert System (MES).

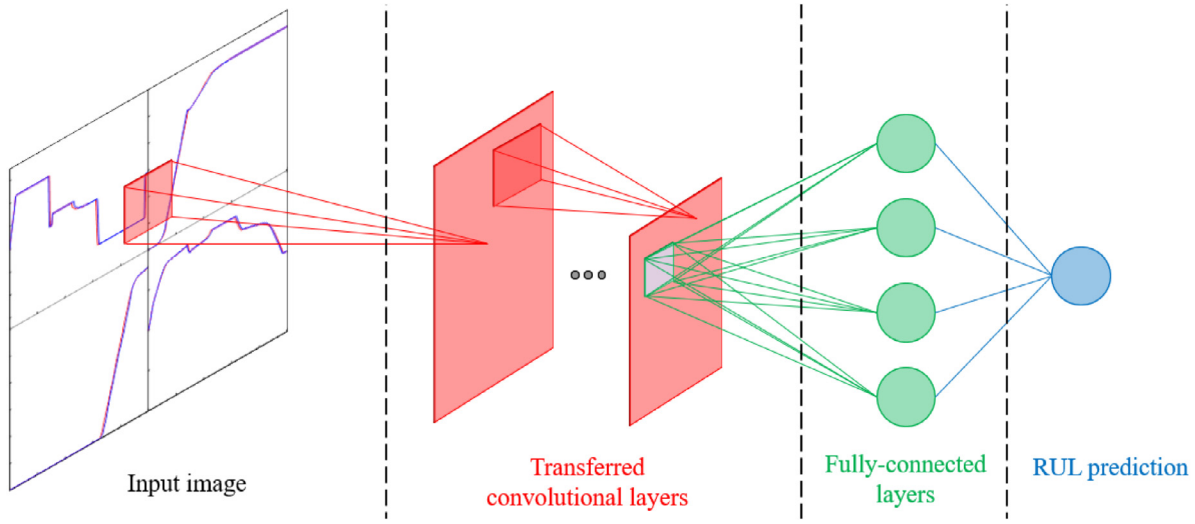

Fig. 5. Workflow example for the transferred neural networks.

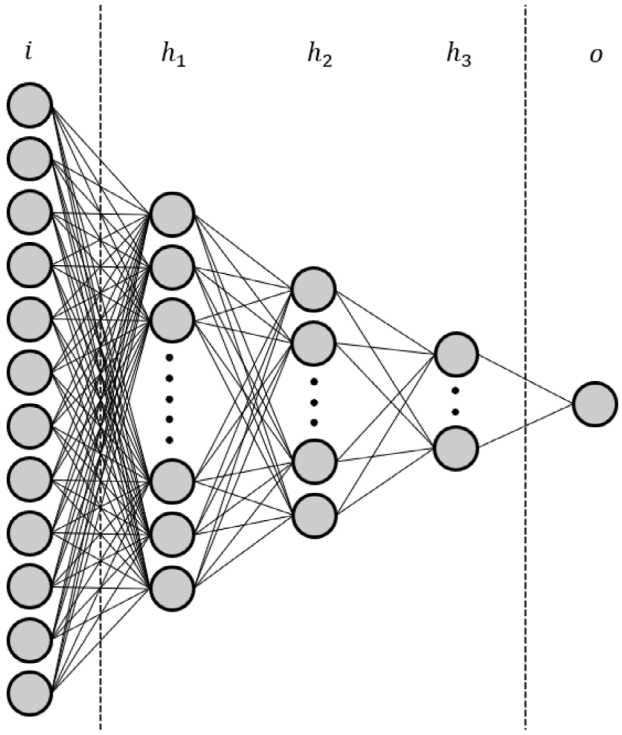

Fig. 6. Fully-connected neural network architecture.

Table 1
Fully-connected neural network's hyperparameters.

Layer	i	h_1	h_2	h_3	o
Number of neurons	12	2048	512	128	1

6.1. Fully-connected neural network

The developed FCNN had the task of acting as a benchmark model before the addition of the image dataset to allow us to quantify the benefits of the transfer learning models, alongside validating the health indicators chosen for this study. The following figure shows the developed architecture of the artificial neural network (see Fig. 6 and Table 1).

Using this architecture alongside the previously outlined inputs, we managed to achieve an 82.05% RUL prediction accuracy on the test set whilst using inputs of 10 consecutive cycles. The following table shows more detailed accuracy metrics for the developed network alongside comparison networks from the literature. Note that the presented FCNN notation within Table 2 corresponds to the number of cycles shown within the inputs.

It is important to note that the Elastic Net model was able to achieve a higher prediction accuracy, but this was at the cost of showing the first 100 cycles of data, also, the RMSE value is quite high, indicating that the method often gives out a large prediction error. We can also see that when compared to the DCNN from Hong et al. (2020), our method achieved very similar results when using only an additional cycle and about half as many training parameters, which validates the network and our health indicators. The FCNN's results will be used as a benchmark to quantify the improvements that can occur by using images alongside the numerical inputs. The FCNN_1 network, which used the single cycle of data, used slightly less trainable parameters due to Qc_d and Qd_d being removed from the health indicators, since no capacity variation occurred, meaning that only 10 health indicators were used as inputs for this specific dataset.

6.2. Transfer learning

This section will compare the performance of different available networks when using only the image datasets. For this, we imported various transferable networks and redefined their fully-connected layers, which act as classifiers. All networks were tested using the same classifying architecture, which can be seen in Fig. 7. Note that not all available networks were trained and compared, instead, we sought a few of the top-performing networks to allow us to provide a proof of concept (see Table 3).

The Tables 4–6 presents the RUL prediction results of various pre-trained networks available within the PyTorch library when trained on the 10, 5, and 1 cycle image datasets.

Here, the number of trainable parameters varies significantly due to the networks' architecture and the final dimensions of their feature map. From these results, we can conclude that if the existing methodology of health indicators in the form of numerical inputs would be replaced, image-based transfer learning networks would be able to accurately predict battery prognostics by themselves. Mainly, the VGG11 network has the highest accuracy performance of 86.86%, representing a 5.6% increase from the FCNN. A few networks were able to outperform the benchmark FCNN that was developed within

Table 2
Results of the FCNNs on the health indicator inputs.

Model	# of trainable parameters	Shown cycles	MAE (cycles)	RMSE (cycles)	Accuracy (%)
Elastic Net (Severson et al., 2019)	N/A	100	N/A	214	89.3
DCNN (Hong et al., 2020)	2,393,468	4	65	N/A	80.3
MES (Sanz-Gorrachategui et al., 2021)	N/A	10	N/A	49	84.8
FCNN_1	1,137,409	1	66.74	108.9	81.26
FCNN_5	1,141,505	5	63.51	102.88	81.98
FCNN_10	1,141,505	10	62.57	109.23	82.05

Table 3
Image network's classifier hyperparameters.

Layer	i	h_1	h_2	o
Number of neurons	Varying	512	128	1

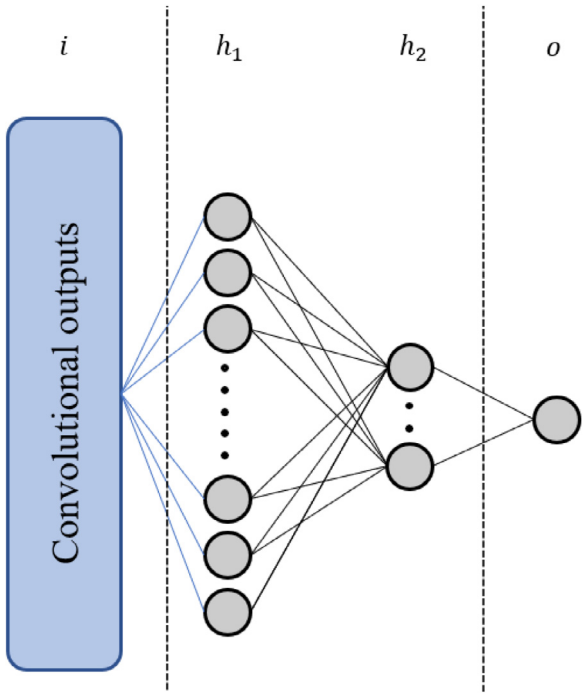

Fig. 7. Image network's classifier architecture throughout the imported models.

this study using only the image of the current, voltage, and capacity curves, indicating a strong ability to learn from the images shown.

It would seem like there is an overall correlation between the number of trainable parameters and the prediction accuracy, although this trend does not hold as true for networks other than the AlexNet and the VGG11. It is also interesting to note that the total number of parameters does not correlate to the overall accuracy, with the EfficientNet and the Resnet152 achieving a lower performance than our benchmark FCNN. Surprisingly, the image-based networks managed to achieve high accuracy even when shown only a single cycle's worth of data. To the best of the author's knowledge, this would make it one of the fastest available RUL prediction methods to date, enabling an accurate prediction using only one charging and discharging cycle.

To better prove the advantages of transfer learning for such an application, a short comparison is made to highlight its benefits. Firstly, the previously obtained results for the AlexNet and VGG11 networks, representing the usage of the neural networks' pre-trained weights, are compared with the weights being randomly initialized. This causes the entire network to be trained on the batteries' images of data curves, which is why the number of trainable parameters will differ whilst the total number of parameters will remain the same for both cases. Note that the overall architecture remains unchanged for this comparison. The 1 cycle dataset is used to obtain the results presented in Table 7 and the average time taken per training epoch was obtained using a NVIDIA RTX 3090 graphics card.

From these results, it can be observed that the networks' accuracy remains similar for the pre-trained and the randomly initialized models. It is expected that this similarity would dissipate if the dataset would be smaller, as the randomly initialized neural networks might not have sufficient data to adequately train their entire architecture. However, this experiment proves that pattern information gathered from training on the ImageNet dataset is indeed relevant for recognizing images of curves and extracting useful information from them. It should also be observed that using the pre-trained models granted a much faster training and testing response when compared to their counterparts. This has great applicability for offline scenarios as the computational hardware might be limited.

Additionally, it should be noted that the pre-processing labor required to generate these complex architectures is completely removed when making use of pre-made models, thus allowing their usage to be had to a much wider public due to requiring significantly less in-depth knowledge to develop efficient machine learning prediction models. Transferring such models allows the everyday users to benefit from the extensive expertise and mainly, the resources, of large technology companies.

6.3. Hybridization

This section combines both previous methodologies into a single transfer learning parallel hybrid (TLPH) network, which is now able to make use of single numerical values by using health indicators and large datasets in the form of images. The proposed architecture for the hybrid network is presented in Fig. 8. There we can see that the transfer learning network's outputs are reduced to 256 before being combined with the HIs, this was because different networks had different output sizes, some going up to 4096 outputs. To allow a fair comparison between the different networks, we opted to reduce this number and keep it consistent throughout the study (see Table 8).

The results for this proposed hybrid architecture are presented in Tables 9–11.

The VGG11 network seems to again have an advantage over the other compared network, followed closely by the AlexNet, this is most likely due to their large number of trainable parameters. Although this theory seems to hold up in this scenario, when re-training the entire transferred network, no significant improvements were observed, which would indicate that the size of the output layer is the driving factor of the networks' accuracy. We can also see clear improvements across the board when comparing the TLPH method with the traditional HI and transfer learning methods. Surprisingly, the 1 cycle dataset achieved the highest accuracy out of the three studied datasets, proving that there is sufficient degradation information within a single cycle if enough training data is used.

6.4. Comparison with other studies

This section briefly compares the results from this study with the other previously mentioned papers that made use of the same dataset to predict the RUL of battery cells. Since the TLPH and transfer learning methods achieved their highest performance whilst making use of the

Table 4
Transfer learning results of the models trained solely on the 10 cycle image dataset.

Model	Total # of parameters	# of trainable parameters	MAE (cycles)	RMSE (cycles)	Accuracy (%)
AlexNet (Krizhevsky, 2014)	7,320,385	4,850,689	57.64	89.21	85.75
Resnet18 (He et al., 2015)	11,570,753	394,241	66.12	101.75	80.85
Resnet50 (He et al., 2015)	24,688,705	1,180,673	69.54	105.57	77.75
Resnet152 (He et al., 2015)	59,324,481	1,180,673	68.42	102.47	82.41
VGG11 (Simonyan and Zisserman, 2015)	22,197,633	12,977,153	54.14	83.83	86.62
GoogleNet (Szegedy et al., 2014)	6,256,289	656,385	64.83	97.22	79.10
EfficientNet (Tan and Le, 2020)	65,229,777	1,442,817	69.55	104.32	79.77
DenseNet (Huang et al., 2018)	27,734,593	1,262,593	61.05	93.36	82.76

Table 5
Transfer learning results of the models trained solely on the 5 cycle image dataset.

Model	Total # of parameters	# of trainable parameters	MAE (cycles)	RMSE (cycles)	Accuracy (%)
AlexNet	7,320,385	4,850,689	57.9	88.08	85.5
Resnet18	11,570,753	394,241	67.61	104.08	80.99
Resnet50	24,688,705	1,180,673	67.89	102.08	78.13
Resnet152	59,324,481	1,180,673	68.08	99.42	76.19
VGG11	22,197,633	12,977,153	55.60	84.89	86.32
GoogleNet	6,256,289	656,385	66.68	100.71	77.61
EfficientNet	65,229,777	1,442,817	70.83	107.44	78.50
DenseNet	27,734,593	1,262,593	60.02	92.03	82.89

Table 6
Transfer learning results of the models trained solely on the single cycle image dataset.

Model	Total # of parameters	# of trainable parameters	MAE (cycles)	RMSE (cycles)	Accuracy (%)
AlexNet	7,320,385	4,850,689	56.81	87.42	86.20
Resnet18	11,570,753	394,241	64.21	98.55	83.32
Resnet50	24,688,705	1,180,673	64.93	97.13	81.84
Resnet152	59,324,481	1,180,673	65.01	97.86	81.81
VGG11	22,197,633	12,977,153	54.99	85.05	86.86
GoogleNet	6,256,289	656,385	65.32	100.33	82.28
EfficientNet	65,229,777	1,442,817	69.63	105.23	80.40
DenseNet	27,734,593	1,262,593	61.15	93.48	84.18

Table 7
Comparison of transfer learning models using pre-trained and randomly initialized weights.

Model	# of trainable parameters	Percentage of the network to train (%)	MAE (cycles)	RMSE (cycles)	Accuracy (%)	Average time taken per training epoch (s)
Pre-trained AlexNet	4,850,689	66.26	56.81	87.42	86.20	29.84
Randomly initialized AlexNet	7,320,385	100.00	60.05	86.77	85.37	34.22
Pre-trained VGG11	12,977,153	58.46	54.99	85.05	86.86	71.03
Randomly initialized VGG11	22,197,633	100.00	55.11	87.00	87.10	127.59

Table 8
TLPH neural network's hyperparameters.

Layer	h_1	h_2	h_3	h_4	o
Number of neurons	256	512	1048	256	1

VGG11 network, this one will be selected for comparison within the following table. The values presented in Table 12 were taken directly from the literature. As mentioned previously, the use of the same dataset throughout these studies will allow an accurate accuracy comparison between the models.

Our proposed methodology outperformed most models currently available in the literature, both in terms of accuracy and of required cycles to obtain an accurate estimation, except for the Elastic Net, which managed to achieve an 89.3% accuracy, although at the cost of a much higher RMSE, with the use of the first 100 cycles. This increased RMSE would suggest that even with a slightly higher MAPE, our proposed model is significantly more robust to different ranges of data than the Elastic Net for this application. Also, since this study aimed to achieve a rapid RUL prediction, we consider the 1.32% trade-off in accuracy to be validated by the 99 cycles spared. Additionally, we reiterate that this study was not seeking to replace existing methods but would instead propose that if the Elastic Net would be used in practical applications, it be hybridized with image inputs to further increase its accuracy.

7. Conclusion

Image-based machine learning was implemented in parallel to the traditional health indicator methodology for use in a regression application where it successfully outperformed popular state-of-the-art networks from the literature, both in terms of accuracy and speed, requiring significantly less input data than the other models. Transfer learning brings the expertise and resources of large technology companies, such as Google and Microsoft to the everyday user, thus allowing regression fields to make use of their advancements. This method is easily implementable, completely open-source, and has potential not only for battery prognostics prediction but also for a variety of other regression tasks. A benchmark network that uses health indicators extracted from battery cycling data was developed to quantitatively assess the improvements to be had by adding images to the list of inputs. Additionally, a proof of concept was presented by transferring

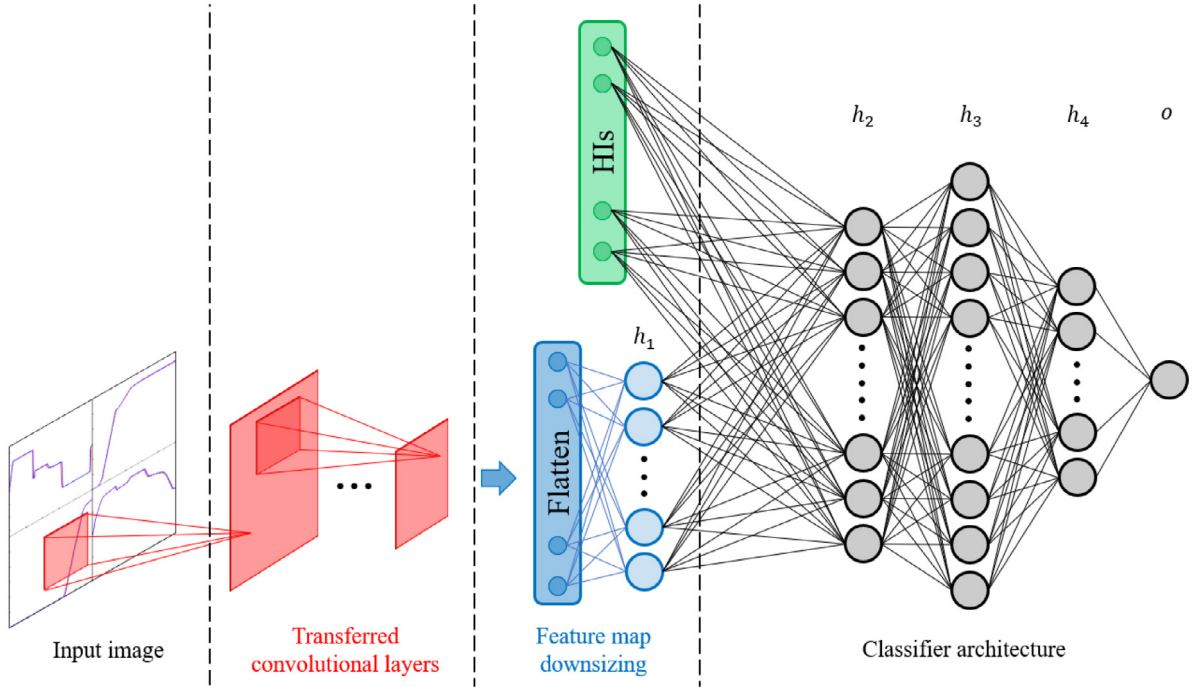

Fig. 8. Proposed TLPH neural network architecture.

Table 9
Transfer learning parallel hybrid with the 10 cycles dataset.

Model	Total # of parameters	# of trainable parameters	MAE (cycles)	RMSE (cycles)	Accuracy (%)
AlexNet	5,773,937	3,304,241	55.83	90.23	85.99
Resnet18	12,252,529	1,076,017	59.60	97.52	85.30
Resnet50	24,977,265	1,469,233	56.28	95.33	85.37
Resnet152	59,613,041	1,469,233	63.24	108.11	83.48
VGG11	16,587,953	7,367,473	51.04	79.71	87.20
GoogleNet	6,806,993	1,207,089	61.17	107.06	83.72
EfficientNet	65,387,265	1,600,305	68.15	131.69	81.30
DenseNet	27,982,193	1,510,193	58.24	99.32	84.73

Table 10
Transfer learning parallel hybrid results with the 5 cycles dataset.

Model	Total # of parameters	# of trainable parameters	MAE (cycles)	RMSE (cycles)	Accuracy (%)
AlexNet	5,773,937	3,304,241	54.16	86.94	86.38
Resnet18	12,252,529	1,076,017	60.27	98.63	84.85
Resnet50	24,977,265	1,469,233	56.75	92.33	84.51
Resnet152	59,613,041	1,469,233	57.38	93.66	84.35
VGG11	16,587,953	7,367,473	50.60	79.07	87.15
GoogleNet	6,806,993	1,207,089	56.75	92.33	84.51
EfficientNet	65,387,265	1,600,305	67.03	111.79	83.27
DenseNet	27,982,193	1,510,193	57.05	88.52	84.68

Table 11
Transfer learning parallel hybrid results with the 1 cycle dataset.

Model	Total # of parameters	# of trainable parameters	MAE (cycles)	RMSE (cycles)	Accuracy (%)
AlexNet	5,772,909	3,303,213	46.03	72.59	87.84
Resnet18	12,251,501	1,074,989	55.06	95.05	86.00
Resnet50	24,976,237	1,468,205	60.54	100.9	85.50
Resnet152	59,612,013	1,468,205	56.07	90.46	85.71
VGG11	16,586,925	7,366,445	47.67	73.35	87.98
GoogleNet	6,805,965	1,206,061	62.55	107.5	83.86
EfficientNet	65,386,237	1,599,277	69.59	121.49	82.08
DenseNet	27.981.165	1.509.165	52.74	83.32	85.64

Table 12
Comparison of this study's result with available methods in the literature.

Model	# of trainable parameters	Shown cycles	MAE (cycles)	RMSE (cycles)	Accuracy (%)
Elastic Net	N/A	100	N/A	214	89.3
DCNN	2,393,468	4	65	N/A	80.3
MES	N/A	10	N/A	49	84.8
Transfer learning	12,977,153	1	54.99	85.05	86.86
TLPH	7,366,445	1	47.67	73.35	87.98

a publicly available pre-trained neural network for the prediction of battery RUL using only images of data curves as inputs. This study proved the benefits of adding image-based inputs for the rapid and accurate prediction of batteries' RUL by improving the prediction accuracy by almost 7% through hybridization, using no additional data. This research also concluded that images can provide an accurate prediction on their own, without the need for hybridization, and without the need to develop a novel neural network architecture with the use of already available transfer learning networks.

8. Future work

Many aspects of this study could be adapted for different applications, whether it be for other battery health prognostics prediction, fault diagnosis, or even for other fields that require the need for non-linear prediction models. The use of images as inputs has broad uses for any function that might utilize data curves. Also, wider comparisons could be made to determine whether other types of machine learning models might be more accurate for image-based inputs used for regression tasks, such as the one conducted in Sharma et al. (2022) for a classification task. Our proposed methodology could also be validated on different battery chemistries, and with a different set of inputs. Some possible improvements to this work to improve the overall prediction accuracy could include more inputs on the image, or by including more curves per input. It would also be possible to alter the proposed hybrid architecture further to be used as a standalone model by implementing LSTM layers before the health indicator inputs to better make use of the historical cycling data, and even CNN layers prior to the LSTMs' to make use of their input filtering abilities as proposed in the model from Hong et al. (2020). Some limitations of the proposed methodology are that although it does not require as much data to train as other machine learning implementations, the quality and volume of the dataset are still driving factors in the overall performance of the models.

CRediT authorship contribution statement

Jonathan Couture: Conceptualization, Methodology, Software, Investigation, Validation, Writing – original draft. **Xianke Lin:** Conceptualization, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

XL acknowledges support from the Natural Science and Engineering Research Council of Canada (NSERC) through the NSERC Early Career Researcher Supplement & Discovery Grant Program (RGPIN-2018-05471) and Ontario Tech University Startup Fund, Canada.

Code availability

Code for the data processing and modeling work is available from any of the authors upon request.

References

Bozinocski, S., Fulgosi, A., 1976. The influence of pattern similarity and transfer of learning upon training of a base perceptron B2. In: Proceedings of Symposium Informatica.

Casal, L.C., Rodriguez, M., Corchero, C., Carillo, R.E., 2019. Evaluation of the end-of-life of electric vehicle batteries according to the state-of-health. World Electr. Veh. J. 10.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep residual learning for image recognition. arXiv.

Hong, J., Lee, D., Jeong, E.-R., Yi, Y., 2020. Towards the switch prediction of the remaining-useful life of lithium-ion batteries with end-to-end deep learning. Appl. Energy 278.

Huang, G., Liu, Z., Maaten, L. van der, Weinberger, K.Q., 2018. Densely connected convolutional networks. arXiv.

Kocer, B., Arslan, A., 2010. Genetic transfer learning. Expert Syst. Appl. 699, 7–7002.

Krizhevsky, A., 2014. One weird trick for parallelizing convolutional neural networks. arXiv.

Li, X., Zhang, L., Wang, Z., Dong, P., 2019. Remaining useful life prediction for lithium-ion batteries based on a hybrid model combining the long short-term memory and Elman neural networks. J. Energy Storage 21, 510–518.

Liu, K., Shang, Y., Ouyang, Q., Widanage, W.D., 2021. A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery. IEEE Trans. Ind. Electron. 68, 3170–3180.

Liu, W., Xu, Y., 2019. A comprehensive review of health indicators of li-ion battery for online state of health estimation. In: 2019 IEEE 3rd Conference on Energy Internet and Energy System Integration, Vol. 120. pp. 3–1208.

Liu, D., Zhou, J., Pan, D., Peng, Y., Peng, X., 2015. Lithium-ion battery remaining useful life estimation with an optimized Relevance Vector Machine algorithm with incremental learning. Measurement 63, 143–151.

Ma, G., Zhang, Y., Cheng, C., Zhou, B., Hu, P., Yuan, Y., 2019. Remaining useful life prediction of lithium-ion batteries based on false nearest neighbors and a hybrid neural network. Appl. Energy 253.

Ren, L., Dong, J., Wang, X., Meng, Z., Zhao, L., Deen, M.J., 2021. A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life. IEEE Trans. Ind. Inf. 17, 3478–3487.

Roy, A.M., 2022. A multi-scale fusion CNN model based on adaptive transfer learning for multi-class MI-classification in BCI system. http://dx.doi.org/10.1101/2022.03.17.481909, bioRxiv.

Saha, B., Goebel, K., 2007. NASA Ames Prognostics Data Repository [WWW Document]. NASA, URL http://ti.arc.nasa.gov/project/prognostic-data-repository.

Sanz-Gorrachategui, I., Pastor-Flores, P., Pajovic, M., Wang, Y., Orlik, Bernal-Ruiz, C., Bono-Nuez, A., Artal-Sevil, J.S., 2021. Remaining useful life estimation for LFP cells in second-life applications. IEEE Trans. Instrum. Meas. 70, 1–10.

Severson, K.A., Attia, P.M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M.H., Aykol, M., Herring, P.K., Fraggedakis, D., Bazant, M.Z., Harris, S.J., Chueh, W.C., Braatz, R.D., 2019. Data-driven prediction of battery cycle life before capacity degradation. Nature Energy 38, 3–391.

Sharma, R., Kim, M., Gupta, A., 2022. Motor imagery classification in brain-machine interface with machine learning algorithms: Classical approach to multi-layer perceptron model. Biomed. Signal Process. Control 71, 103101. http://dx.doi.org/10.1016/j.bspc.2021.103101.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. arXiv.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., 2014. Going deeper with convolutions. arXiv.

Tan, M., Le, Q.v., 2020. Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv.

Zhang, Y., Xiong, R., He, H., Pecht, M.G., 2018. Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. IEEE Trans. Veh. Technol. 67, 5695–5705.

Zhou, Y., Huang, M., 2016. A novel health indicator for on-line lithium-ion batteries remaining useful life prediction. J. Power Sources 321, 1–10.

Zraibi, B., Okar, C., Chaoui, H., Mansouri, M., 2021. Remaining useful life assessment for lithium-ion batteries using CNN-LSTM-DNN hybrid method. IEEE Trans. Veh. Technol. 70, 4252–4261.