Real-Time Estimation of Bath Temperature in Aluminium Electrolysis Cells Using Temporal Neural Networks

Alexandre Beaulieu[1], Diego Salomone[2], David Munoz[3], Bertrand Allano[4] and Mohamad Ismail[5]
1. Data Scientist
2. Lead Data Scientist
Rio Tinto – Industry 4.0, Saguenay, Canada
3. Experienced Project Manager
4. Reduction Principal Advisor, Rio Expert
Rio Tinto – Solutions Technologies Aluminium, Voreppe, France
5. Modelling Engineer
Rio Tinto – Solutions Technologies Aluminium – LRF, Saint-Jean-de-Maurienne, France
Corresponding author: Alexandre.Beaulieu2@riotinto.com
https://doi.org/10.71659/icsoba2024-al036

Abstract

The bath temperature is a critical parameter to monitor in aluminium electrolysis cells as it significantly influences various phenomena within the cell, including side ledge and bottom freeze build-up, which can impact cell life. In today's energy landscape, there is growing pressure to modulate power input in potlines. This pressure is driven by the increasing reliance on renewable energy sources, which is often cyclical in nature, and from energy providers who wants to offset energy consumption peaks. Power modulations have a substantial influence on the bath temperature. Accurate and real-time bath temperature estimation hence becomes essential to monitor and manage optimal cell productivity conditions.

In this paper, we propose a virtual sensor based on Temporal Convolutional Neural Networks (TCN) to estimate bath temperatures in real-time. The proposed solution achieves a Mean Absolute Error (MAE) of approximatively 3 degrees Celsius, while never using any measurement of bath temperature as input. This level of accuracy is essential for effective power modulation and maintaining optimal bath conditions. It also allows for time-efficient scenario evaluation, leveraging a better decision making around modulations.

Our work highlights the potential of advanced Machine Learning (ML) techniques to perform real-time, accurate estimation of bath temperatures. By demonstrating the feasibility of this approach, we aim to pave the way for more modern aluminium electrolysis processes.

Keywords: Artificial Intelligence, Real-time, Bath temperature, Virtual sensor.

1. Introduction

Bath temperature is a critical parameter in the aluminium reduction process. Optimal aluminium generation occurs within a narrow temperature range between 960 °C and 970 °C. When the bath temperature is outside this thermal operational window, the metal productivity is reduced. Other than current efficiency, higher temperatures can reduce the lifespan of the cell by melting the side ledge around the cell and exposing the cell sides to lining erosion. On the other hand, lower temperatures can cause the superfluous bottom freeze that can cause uneven cathodic current distribution and uneven wear of the carbon blocks. In the industrial context, the monitoring of the temperature is done by manually inserting a thermocouple in the bath. This labor-intensive process is typically conducted once per day.

The energy landscape is undergoing a rapid transformation. The energy infrastructure is being put to the test by growing demands both from private consumers and the industrial sector. The increase in development of renewable energy sources such as wind and solar is changing the availability of energy at a given time due to their cyclical nature. [1] This phenomenon is exacerbated as energy demands vary during the day, corresponding to peak demand in households in the morning and evening, and through the year as seasonal temperatures fluctuate (heating during cold winter days and climatization during heat waves). Power companies are starting to create incentives to encourage the consumers to offset energy consumption during peak demand periods. One of these incentives is modulating the prices, lowering it during low-demand periods and increasing it during high-demand periods.

In the aluminium industry, power modulations offer a strategic approach to minimize energy-related expenditures. The goal is to reduce power consumption by decreasing the potroom current intensity during peak consumption periods, where prices are at their highest, and take advantage of off-peak periods by increasing the intensity. Since aluminium smelting is inherently energy-intensive, optimizing energy consumption by modulating the potline intensity can result in substantial cost savings. However, modulations can have a significant impact on the thermal balance of the cells. During lowered intensity, cells will cool, and heat-up during high intensity periods, creating substantial changes in bath temperature. Reaching temperatures too far out of the thermal operational target can have devastating consequences for cell performance and longevity.

To ensure that the bath temperatures of the cells will remain within an acceptable range before applying modulations, different algorithms can be used to estimate the behaviour of the bath temperatures during modulations. One approach is to create a physical model. A physical model is composed of mathematical equations that describe the thermal dynamics of the cells, including heat transfer, conduction, and radiation. In the case of the aluminium reduction process, a physical model would incorporate factors such as the cell's geometry, material properties, and operating conditions to accurately predict temperature fluctuations. The main advantage of this approach is that if the model is complex enough (encodes enough physical interactions), it can give estimates close to the reality. On the other hand, these algorithms are quite computationally expensive. Depending on the model, it can take more time to simulate than the simulation horizon, rendering it hard to use in real time, or at scale.

While using physical models provide more explainable results, using neural networks is also a viable approach. Neural networks are a type of machine learning model inspired by the structure and function of the human brain. At their core, neural networks consist of layers of interconnected nodes (neurons) that process and transmit information. One of the key advantages of neural networks is their ability to learn from large amounts of data, uncovering underlying patterns and relationships that might be difficult or impossible for humans to discern on their own. Despite requiring more learning time, once trained, neural networks can operate at high speeds, making them suitable for real-time applications. Moreover, advancements in neural network architectures have led to the development of specialized models tailored for processing time series data [2]. In this study, we employ the Temporal Convolutional Neural Network (TCN) architecture [3], which has been shown to excel on key benchmarks since its introduction in 2018, surpassing established techniques such as LSTM (Long-Short-Term Memory) networks [4], previously considered industry standards.

This paper aims to demonstrate the potential of neural networks for real-time estimation of bath temperature. For the context of this paper, the threshold of a reasonable model for bath temperature prediction was considered to be below 5 °C.

2. Proposed Solution

In the context of power modulation, a valid solution would be to estimate the direct impact of the power modulation on the cell's bath temperature. In contrast to this approach, our solution aims to estimate the real-time temperature of the bath during the modulation process. This approach offers several benefits over simply estimating the final temperature.

Firstly, there is not a lot of data available in the deep learning sense. Deep learning models are often trained with millions, billions, and even trillions of examples. It is unrealistic to produce millions of modulations instances for a model to learn from. To fix this lack of scale on the learning data, the modulation period was broken into smaller time intervals. This allowed for more granularity of available data related to the thermal balance of the cell and its interactions with various cell events. With this the model could be able to capture the complex relationships between inputs and outcomes, rather than viewing a single modulation event as a standalone prediction.

Secondly, our approach is more versatile. The foremost application is exploiting the model as a virtual sensor by providing more frequent insight into the bath temperature of the cells. This means that the model can also be useful outside of modulation periods.

2.1 Model

The proposed solution leverages the TCN architecture. TCN is a neural network architecture designed specifically for sequential data. TCN's unique strengths stem from the following three key characteristics:

Causal convolution. TCN uses a fully convolutional networks architecture with zero padding and causal convolution. [5] A fully convolutional architecture keeps hidden layers at the same size as input layer. The use of causal convolutions in TCN is particularly noteworthy. A convolution is causal when an output at time *t* is convolved only with elements from time *t* and earlier in the previous layer. In other words, any output is based exclusively on previous data.

Dilated convolution. [6] Dilated convolutions introduce gaps between filter elements, allowing the filter to capture larger regions of the input without increasing the number of parameters. The dilation is increased in proportion to the dilation of the previous layer's dilation. The resulting model can extract temporal patterns at different scales, while keeping a small footprint. As illustrated in Figure 1, dilated convolutions yield a unique connectivity pattern, where each filter element skips over specified intervals, effectively expanding its receptive field without introducing additional parameters.

Residual Connections. [7] A residual connection is shortcut that connects the input and output of a layer or a block (or group of layers). This allows the network to learn more complex representations without having to worry about vanishing gradients, which can occur when using traditional feedforward networks with many layers. A residual block combines the inputs of block to the output of the block. This approach has shown benefit in very deep neural networks.

TCN have shown impressive performance on multiple benchmark datasets, outperforming more conventional models such as LSTM. A notable example is achieving state of the art performance (97.2 prediction accuracy, versus 85.7 for LSTM) on the Permutated Sequential MNIST dataset. MNIST [8] is one of the most well-known datasets in machine learning. It consists of black and white scans (28 × 28 pixels) of handwritten digits from 0 to 9. A variation of this dataset is the sequential version, where images are transformed into a sequence of pixels (784 × 1). The permutated version is a further variation where the order of the sequence is randomly permuted. Good performances on this dataset show the ability of a model to consider valuable information spread randomly across sequences. (see [9] for a comparison of some model performances on the permutated sequence MNIST dataset).

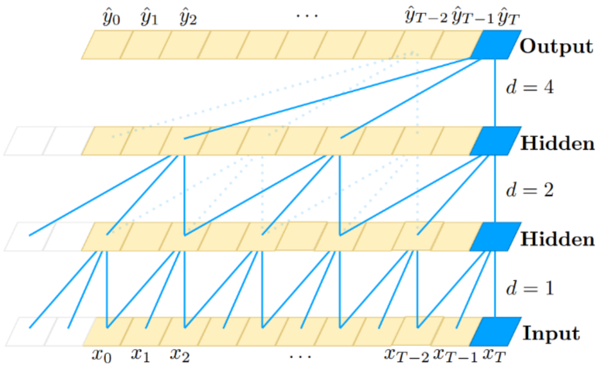

Figure 1. A dilated causal convolution with dilation, a key element of the TCN architecture, taken from [3].

2.2 Data

For this work, we used measurements performed on two prototype cells installed in the Laboratoire de Recherche des Fabrications (LRF) of Rio Tinto, in Saint-Jean-de-Maurienne. Contrary to cells in industrial smelters, the prototype cells were equipped with a thermocouple providing real-time (every minute) bath temperature measurements. The cells operate at 500 kA. It is important to note that power modulation experiments were actively being performed on the cell for the duration of available measurements. This results in temperatures and cell conditions that varies rapidly and strongly in time.

For training and testing, data is available as far back as January 2022 and up to December 2023. Not all recorded bath temperatures were kept in this period. To ensure proper training, periods where there was cell shutdown, and where measures were invalid or were atypical, were dropped. During the power modulation trials, only the recorded bath temperatures between 900 ºC and 1000 ºC were kept, ensuring that more data was available to train. Also, this ensured that the dataset contained instances outside the optimal range of operation. In total, there is an equivalent of 284 days (about 9 and a half months) of continuous data available.

The actual modulations, their durations, frequencies, and intensities are varied and were part of a different experiment. This heterogeneity is interesting since it provides a variety of combinations for the training routine.

To train and validate the model, the available data was split. The month of November 2023 was set aside for the testing set. This month contained a continuous modulation experiment. This month was selected as it is the expected final mode of operation for the model. The test set contains the equivalent of 12 days of continuous measurements.

2.3 Inputs

The different model inputs come from multiple sources, some real-time timeseries and other manual samples coming from traditional cell operations available daily. Some features are derived from the aforementioned sources. On the real-time inputs available, we exploit the cell resistance (µΩ), the target resistance (µΩ), the current (A), feeding period (s), low and high frequency noise in the cell resistance, instability, and additional resistance (μΩ), anode change event, metal tapping event, anode cover event and anode effects. From the available resistances R and current, I, we directly compute the power (RI^2). The events on the cells are provided as binary flag (indicating that the event is occurring now) and an integer indicating the time since the last event occurred. Further information on the events is computed, such as the anode number of the anode change, and the quantity of metal tapped.

From the analysis and measurements taken on the cells, we used the bath height (cm), metal height (cm), and concentrations (% mass) of AlF_3, Al_2O_3 and CaF_2 in the bath.

It is important to note that there is no temperature as input to the model. This means that the only way that the model can estimate the temperature is by analysing the evolution of other inputs in time.

It is also worth noting that many features impacting the thermal balance of the cells are not available to the model. Temperature information generally available on site, such as the status of Forced Connection Networks (FCN) that provides cooling to the cell, the status of the Gas Treatment Center (GTC), which allows air circulation in the cell and outside temperature are examples of features that could be used in future versions of the model.

2.4 Preparation

Since not all input data is available at the same frequency, we resampled and aligned the data at a frequency of 5 minutes, striking a balance between resolution, model complexity, and target inference rate. While lower-frequency sampling may conceal high-frequency patterns, it reduces the final footprint of the model. Furthermore, we prioritize maintaining a resampling frequency that is equal to or greater than our production prediction rate.

The time series data is then divided into 24-hour windows, resulting in a final window shape of 288×31 (288 5-minute samples and 31 features). This allows us to capture meaningful patterns and trends within the data.

Before feeding the prepared data into our model, we apply two additional processing steps: clipping values within a reasonable range to prevent outliers from affecting predictions, and applying min-max scaling to ensure all features are suitable for the deep learning model.

2.5 Performances

The model demonstrated impressive performance, achieving a Mean Absolute Error (MAE) of 3.5 °C on the training data and a remarkable 2.4 °C on the test set (unseen data). Figure 2 shows the temperature predictions in the test set.

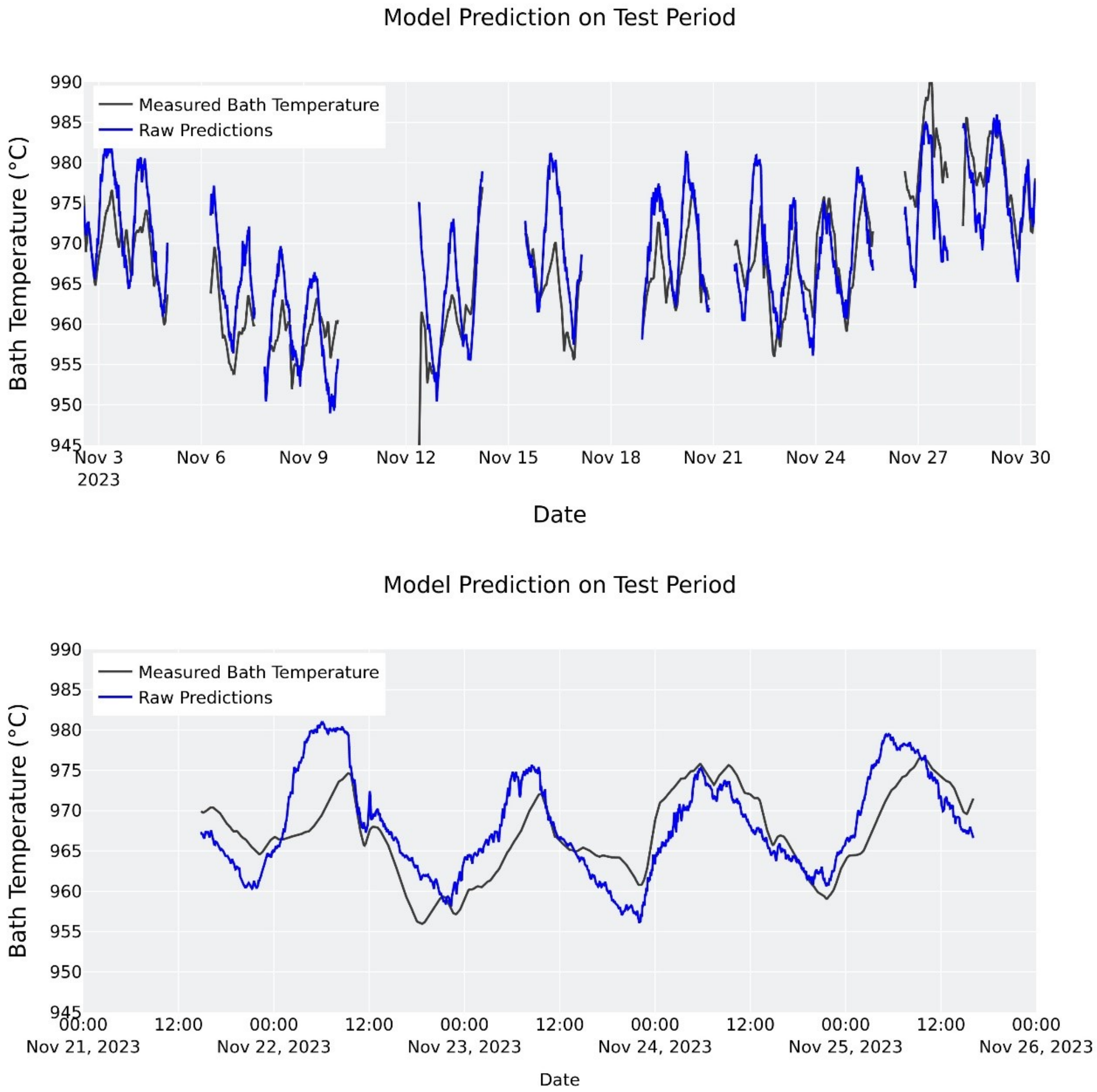

Figure 2. Predictions of the model on the test dataset. Top: whole prediction period. Bottom: few days of consecutive predictions.

This significant reduction in error between the two datasets suggests that the model was able to effectively generalize its knowledge to new, unseen data while maintaining high accuracy on the familiar training set. The relatively small MAE on both sets indicates that the model was highly effective at capturing subtle patterns and relationships in the underlying data, resulting in predictions that were consistently close to the true values, even though no temperature was given as input.

While the actual training of the model can be time consuming, the actual inference time of the model is fast. During tests, the model inferred at a rate of 69 ms per batch of 8 predictions. This means that it can reasonably be used in a real time.

3. Application

Since the model was developed, an application was made to validate the model in a real-time scenario. The application is a dashboard that shows predictions compared to available temperatures measured in real time, and the model predicts the latest temperature every 10 minutes.

The validation period of the model starts at the last available training data (December 2023). For this paper, predictions were made from January 2024 to March 2024 on one prototype cell only. This resulted in 90 days of predictions. This period consists of further power modulation experiments. The experiments were different from the experiments made during the training period and hence represent completely new data. This is interesting as it allows us to appreciate the ability of the model the generalize.

The main difference between the application and the original model is that the predictions are synchronized once per day with the measured temperature. Synchronising the temperatures will be the model's mode of operation in real life (at potroom scale), as it should compensate for the model's lack of temperature context (no temperature is used as input) and potential data drifts. The model is good at guessing changes in temperatures, but not at estimating the base temperature of the cell. This is visible in Figure 3. In this figure we can see that for the shown period, synced temperatures are close to measured temperatures, but raw predictions show a constant gap of approximatively 15 °C below the measured temperature.

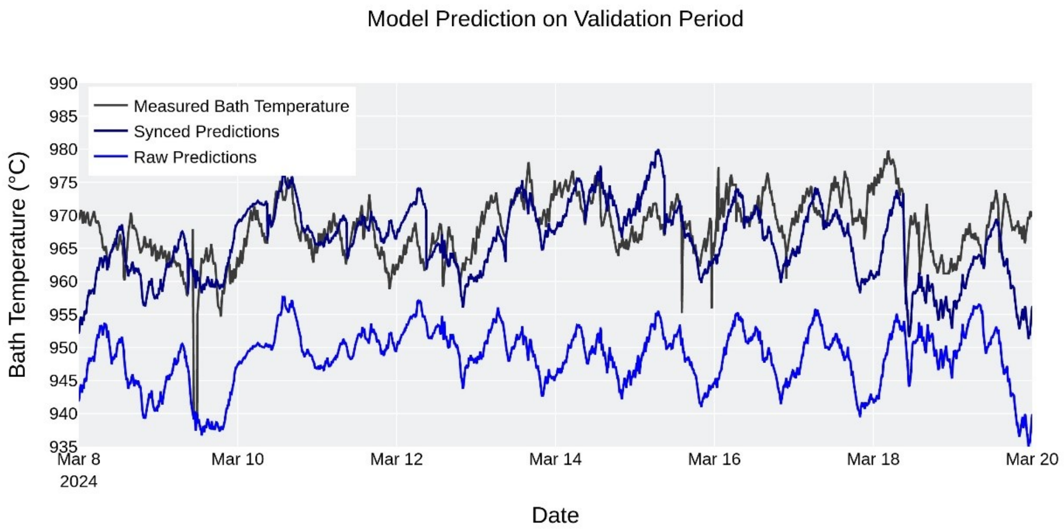

Figure 3. Predictions of the model on the validation period.

The MAE for the whole validation period (since January 2024) is 4.7 °C. Figure 4 shows the MAE of the model per day for the whole validation period. For 67 % of days in the validation period the MAE is below the 5 °C threshold.

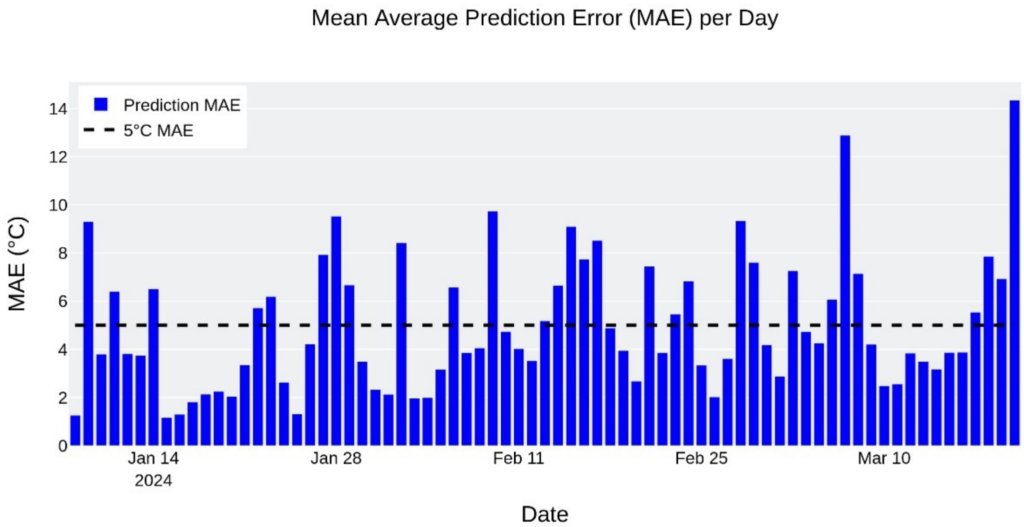

Figure 4. Mean Average Error of the model since application is live, agglomerated per day.

There are multiple potential reasons for this decrease in performance. For starters, since different and new experimentations continued to be made during this period, predictions were made in conditions that were not necessarily seen in the past. While not all variations were explained, some were due to new unseen behaviour in cell operation caused by the extreme impact of the modulations on the bath temperature. Moreover, the validation period does not contain the same proportion of modulation experimentations and normal operations almost 45 % of the validation period consist of normal operation compared to 25 % in the training period. The model shows a propension to over-estimate variations in temperatures in non-modulation periods, which could be explained by this imbalance.

4. Conclusion and Future Works

In this paper, we have demonstrated the effectiveness of using Temporal Convolutional Neural Networks (TCN) as a virtual sensor to estimate bath temperatures in real-time. The model achieved an impressive mean absolute error of 3.5 °C on the training set, and 2.4 °C on the testing dataset, without requiring any direct measurement of bath temperature. The ability of this model was further validated in an application estimating the temperature in real-time, achieving global performances below the threshold of 5 °C of error. We have shown that this approach can provide accurate and timely insights for power modulation.

While in this first phase we were able to demonstrate the bath temperature estimation, the next phase will concern the application of the model in existing and generated scenarios. This would allow us to test whether the model can act as a more generic estimator, much like a physical model. One of the main challenges in this next phase will be to conciliate real historic data and generated data. While generating data is feasible, it is not as easy as it seems. It is important to keep it mind that what can be generated, will seem adequate but will not represent the reality, which can result in bad predictions. Some inputs will be easier to provide, such as the power intensity and target resistance and planned events, which are all known in advance. Others will require more effort to simulate.

Another aspect that can be worked on is the actual model and selected inputs. While TCN has shown state of the art results on many datasets, the last few years have brought a load of new algorithms with their own twist on sequence problems, and often outperforming TCN.

Finally, further work will be needed to access the ability of the model to generalize between a variety of cell designs and technologies.

5. References

1. International Energy Agency (IEA), *World Energy Outlook 2022*, International Energy Agency, https://www.iea.org/reports/world-energy-outlook-2022 (Accessed on 23 May 2024).
2. Bryan Lim and Zohren Stefan, Time-series forecasting with deep learning: a survey, 15 February 2021, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* vol. 379, no. 2194, https://doi.org/10.1098/rsta.2020.0209
3. Shaojie Bai, J. Zico Kolter, and Vladlen Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, arXiv:1803.01271v2 [cs.LG] 19 Apr 2018, https://arxiv.org/pdf/1803.01271 (accessed on 29 July 2024).
4. Sepp Hochreiter and Jürgen Schmidhuber, Long short-term memory, *Neural Computation*, 1997, vol. 9, 1735–1780.
5. Jonathan Long, Evan Shelhamer and Trevor Darrell, Fully convolutional networks for semantic segmentation, 2014, https://arxiv.org/pdf/1411.4038 (accessed on 29 July 2024).
6. Aaron van den Oord et al., WaveNet: a generative model for raw audio, 2016, https://arxiv.org/pdf/1609.03499 (accessed on 29 July 2024).
7. Kaiming He et al., Deep residual learning for image recognition, 2015, https://arxiv.org/pdf/1512.03385 (accessed on 29 July 2024).
8. Yann Lecun et al., Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, 1998, vol. 86, no. 11, 2278–2324.
9. Sequential Image Classification on Sequential MNIST, *Papers with Code*, https://paperswithcode.com/sota/sequential-image-classification-on-sequential (Accessed on 23 May 2024).