Nuclear mass predictions based on convolutional neural networks

Author: David Morales De León
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: Arnau Rios Huguet

Abstract: The precise determination of nuclear masses is essential for understanding atomic nuclei and for applications in astrophysics and nuclear energy. Traditional models like the liquid drop model, with a root mean squared error of $\sigma = 3.94$ MeV, fail to meet the accuracy of 100 keV required for nuclear astrophysics research. This work introduces a novel approach by implementing a convolutional neural network (CNN) and leveraging the spatial structure of the nuclide chart. Two models, I3 and I4, are trained and tested on the AME2016 database, achieving values of $\sigma = 0.67$ MeV and $\sigma = 0.49$ MeV, respectively. Extrapolating to the new nuclei of the AME2020 database, they hold values of $\sigma = 0.64$ MeV and $\sigma = 0.57$ MeV, demonstrating strong generalization capabilities and proving that CNNs constitute a powerful tool for accurate nuclear mass predictions.

Keywords: Nuclear masses, liquid drop model, convolutional neural networks, atomic nuclei.

I. INTRODUCTION

The accurate determination of nuclear masses is of the utmost importance, as it is a fundamental property that characterizes atomic nuclei. This has many implications in fields such as astrophysics and nuclear energy, enabling a deeper understanding of astrophysical processes and a more efficient energy production. One noticeable example is the study of the rapid neutron-capture process (r-process), which demands an accuracy of 100 keV [1] in order to precisely explain the origin of many elements in the universe. Another relevant research field is the synthesis of superheavy nuclei [2], where the nuclear masses are used to calculate the reaction Q-values, the neutron separation energies and the α-decay energies. Consequently, errors in nuclear masses propagate into these quantities, leading to high uncertainties. Hence, it is essential to develop models that can achieve more accurate predictions.

In order to address this challenge, several theoretical models have been proposed. One prominent example is the liquid drop model (LDM) [3], treating the nucleus of an atom as an incompressible nuclear fluid of constant nuclear density. A more sophisticated model based on the previous one is the finite-range droplet model (FRDM) [4], integrating nuclear shell effects, pairing energies and deformations into the LDM. Notwithstanding, these two models are outperformed by the Weizsäcker–Skyrme 4 model (WS4) [5], which stands out for being one of the most accurate theoretical approaches to date. The latter includes a correction for surface diffuseness, a relevant property near the drip lines that describes how the edges of nuclei spread out.

On the experimental side, the Atomic Mass Evaluation (AME) is a comprehensive database that compiles the latest experimental masses of the isotopes known at the moment. It also incorporates theoretical adjustments when the experimental data is limited by precision or when it is not experimentally accessible. The database includes the information of the proton numbers (Z), the neutron numbers (N) and the nuclear binding energies (B).

Recently, Machine Learning (ML) has emerged as a practical tool for learning complex patterns in data and making predictions based on what has been learnt. In this context, different models have proven to reproduce the behaviour of the experimental nuclear masses computed from the AME database. For instance, techniques such as support vector regression (SVR) and Gaussian process regression (GPR) [6] use a physics-based feature space and achieve results that are comparable to traditional models. Moreover, recent research in probabilistic mixture density networks (MDN) [7] has proven to reach precise results as well.

Nevertheless, despite advancements in both theoretical approaches and ML methods, there still remains the need to develop an ultimate framework capable of accurately predicting nuclear masses and reaching an accuracy of 100 keV required for r-process research. In view of this, the aim of the project is to build a model that accurately predicts nuclear masses by implementing a convolutional neural network (CNN) [8], an architecture that is able to take advantage of the spatial structure in which nuclear masses are arranged on the nuclide chart. This project has been inspired by the work of Ref. [9].

II. NUCLEAR MASS

The atomic mass is defined as the total mass of an atom, including the mass of the protons, neutrons, electrons, and the contribution from the nuclear binding energy. The latter refers to the amount of energy required to divide the nucleus into its individual nucleons. Using natural units ($c = 1$), the atomic mass is given by

$$M_A(Z, N) = ZM(^1H) + NM_n - B, \qquad (1)$$

where $M(^1H) = 938.78$ MeV stands for the hydrogen atomic mass and $M_n = 939.57$ MeV for the neutron mass. Their precise values can be found in Ref. [10].

TABLE I. σ values for the theoretical and ML models.

Model	LDM	FRDM	WS4	SVR	GPR	MDN
σ (MeV)	3.94	0.77	0.47	0.39	0.26	0.58

The nuclear mass, M_N, is defined as the mass of a nucleus, computed by subtracting the contribution of the electrons to the atomic mass. This means that both their mass and their electron binding energy must be taken into account:

$$M_N(Z, N) = M_A(Z, N) - [Zm_e - B_e(Z)]\,, \tag{2}$$

where $m_e = 0.5110$ MeV is the mass of an electron and $B_e(Z)$ is the electron binding energy. $B_e(Z)$ is estimated by the formula of Lunney, Pearson and Thibault [11], expressed in eV as

$$B_e(Z) = 14.4381\, Z^{2.39} + 1.55468 \cdot 10^{-6}\, Z^{5.35}. \tag{3}$$

To characterize the precision of a model at predicting nuclear masses, here we employ the root mean squared error, σ. This metric evaluates the difference between the experimental nuclear masses, M_N^{exp}, and the predicted ones, M_N^{pred}. It can be calculated as

$$\sigma = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(\Delta)^2}, \tag{4}$$

where K is the number of nuclei that are being sampled and Δ is defined as

$$\Delta = M_N^{\mathrm{exp}} - M_N^{\mathrm{pred}}. \tag{5}$$

The σ values for the models mentioned in Section I are given in Table I in order of appearance.

III. LIQUID DROP MODEL

As mentioned in Section I, one of the most commonly used nuclear models is the LDM. For this reason, it serves as a reference to compare with the results obtained in the present work. The model is formalized through the Bethe–Weizsäcker semi-empirical mass formula (SEMF):

$$\begin{aligned} B(A, Z) &= a_V A - a_S A^{2/3} - a_C \frac{Z^2}{A^{1/3}} \\ &\quad - a_A \frac{(A-2Z)^2}{A} - \frac{a_P}{A^{1/2}}\delta, \end{aligned} \tag{6}$$

where $A = Z + N$ denotes the mass number, δ is the pairing term,

$$\delta = \frac{(-1)^N + (-1)^Z}{2}, \tag{7}$$

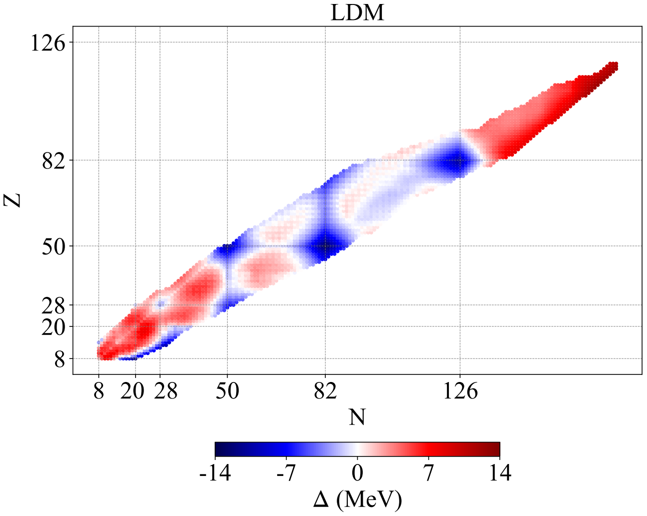


FIG. 1. Scatter plot of Δ (5) between the AME2016 database and the LDM.

and the coefficients a_i can be fitted by using different techniques. For example, the authors of Ref. [12] fitted them to the AME2016 database [10] using the least squares method. In Fig. 1 we show the corresponding nuclide chart, where the x-axis represents the neutron number and the y-axis represents the proton number. Specific neutron and proton numbers are highlighted with grey lines along both axes. They represent the magic numbers [13], specific numbers of nucleons that result in more stable nuclei due to the completion of nuclear shells. A colour bar is used to depict the differences Δ (5) between the experimental values and those predicted by the LDM, ranging from -14 to 14 MeV. Blue indicates an overestimation of M_N while red represents an underestimation. Additionally, each point of the nuclide chart corresponds to a nucleus of the AME2016. Ideally, the LDM would perform perfectly, and the colours would tend to be white, indicating no difference between the experimental masses and the predicted ones. However, as shown in the chart, the LDM does not properly model all the experimental data. For instance, it overestimates the nuclear masses around the magic numbers and underestimates them in the regions of light and superheavy nuclei, i.e. the prominent areas in red.

IV. CONVOLUTIONAL NEURAL NETWORKS

CNNs are a type of neural networks capable of evaluating structured data placed in matrices, like images or grids. They take advantage of spatial arrangements and focus on learning local patterns within the neighbourhood of each element. Networks of this type run by following a two-dimensional convolutional operation, involving the sliding of a small matrix, called kernel, across the grid. The kernel contains the weights, which are the trainable parameters responsible for capturing local features. After the kernel has slid over the entire grid, a feature map is produced, highlighting the most relevant features. This mathematical operation is expressed as

$$F(u,\, v) = \sum_{i=1}^{3} \sum_{j=1}^{3} k(i,\, j) \cdot f(u-i,\, v-j), \qquad (8)$$

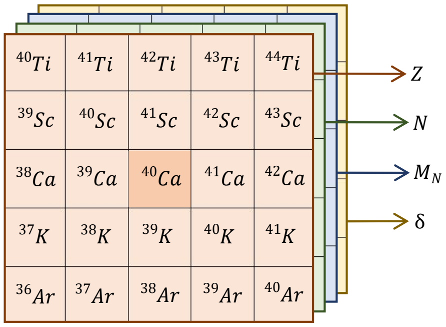
where $F(u,\, v)$ is the value of the element at coordinates $(u,\, v)$ of the output feature map, $k(i,\, j)$ is the kernel at position $(i,\, j)$ and $f(u-i,\, v-j)$ refers to the element of the input matrix at position $(u-i,\, v-j)$. The sums $\sum_{i=1}^{3}$ and $\sum_{j=1}^{3}$ are taken over the dimensions of the kernel, which, in the present work, is a 3×3 matrix. In addition, CNNs also use biases, scalar numbers added after the convolutional operations. They provide more flexibility into the network, as they help the model to adapt better to the patterns of the grid.

In this project, two convolutional layers are used. After performing the convolutions, the outputs undergo an activation function called rectified linear unit (ReLu). This function transforms all the negative values to zero, while the positive values remain intact. By doing so, we introduce non-linearity to the model, enabling it to capture more complex relationships within the grid. Finally, a fully connected layer combines the spatial features that have been extracted by the two convolutional layers and makes a prediction of the nuclear mass for a particular nucleus.

V. MODELS CONSTRUCTION

In the present work, we select the nuclei with $Z \geq 8$ and $N \geq 8$ from AME2016 [10] and AME2020 [14], including 3336 and 3456 nuclei, respectively.

Two models, named I3 and I4, are constructed from scratch using `PyTorch`, following the methodology explained in Section IV. The code details can be found in Ref. [15]. For I3, the input for each nucleus is a $3 \times 5 \times 5$ tensor, meaning that three matrices (or channels) of 5×5 elements are stacked, similar to how an RGB image is built. The three channels correspond to Z, N and M_N. Both Z and N are retrieved from AME2016, while M_N is computed from B, also retrieved from AME2016, using Eqs. (1) and (2). The tensor is centred on the target nucleus that is trying to be predicted and the values surrounding it correspond to its neighbours in the nuclide chart. If one of them is missing in the database, the mean value of M_N from the available neighbours is computed. Even though this method might use hypothetical mass values assigned to nuclei that do not exist in the dataset, by doing so we preserve the spatial structure of the input tensor and allow the CNN to correctly operate. Additionally, the centre of the grid, corresponding to the target nucleus, is set to zero to prevent the model from being biased by the real value.



^{40}Ti	^{41}Ti	^{42}Ti	^{43}Ti	^{44}Ti
^{39}Sc	^{40}Sc	^{41}Sc	^{42}Sc	^{43}Sc
^{38}Ca	^{39}Ca	^{40}Ca	^{41}Ca	^{42}Ca
^{37}K	^{38}K	^{39}K	^{40}K	^{41}K
^{36}Ar	^{37}Ar	^{38}Ar	^{39}Ar	^{40}Ar

FIG. 2. Graphical representation of the construction of the I4 model. Each stacked matrix contains the Z, N, M_N and δ value of the corresponding element, respectively.

In the case of I4, the model is similar to I3, except that it includes a fourth channel, resulting in a $4 \times 5 \times 5$ input tensor. The additional channel represents the nuclear paring effect, δ (7), indicating that pairs of nucleons tend to be more stable when they are coupled rather than when they are alone. This incorporation enhances the model's input information, enabling the construction of a more complex architecture. In Fig. 2 we depict a graphical representation of how the I4 model is built to predict the nuclear mass of the element ^{40}Ca, the target nucleus that is placed in the centre.

Once the 5×5 neighbourhoods are constructed for every nucleus within AME2016, the data is split into training and testing subsets, with 70% used for the training phase and 30% for the testing phase, while making sure that the points are uniformly distributed across the nuclide chart. During training, both the weights and the biases are adjusted to minimize σ of the training subset. Afterward, the model's performance is evaluated by calculating σ for the testing subset (unseen nuclei during training) to assess its generalization capabilities. The entire process is repeated several times, with each iteration being called an epoch. The training process uses backpropagation, an algorithm that computes the gradient of σ with respect to each weight and bias. The gradients indicate how much each parameter needs to be adjusted to minimize σ. For this optimization, we employ the `Adamax` algorithm [16]. Furthermore, both models are trained with early stopping, meaning that after a predefined number of epochs, in which the testing σ does not improve, the training process is stopped. This technique prevents overfitting, i.e. the models fit the training data so close that fail to generalize to unseen data. As a result, the models are trained for approximately 200000 epochs.

VI. RESULTS

The values of σ for both the training and testing subsets are presented in Table II. As shown, the results are slightly superior in I4 than in I3, due to the introduction of the pairing effect. In fact, the I4 model showcases

TABLE II. Results of σ obtained for I3 and I4.

Model	I3		I4	
Phase	Training	Testing	Training	Testing
σ (MeV)	0.36	0.67	0.31	0.49

an improvement of approximately 14% in training and 27% in testing compared to I3. This indicates that the I4 model performs better overall, especially in terms of generalization capabilities. Furthermore, the testing σ values obtained are significantly lower than the σ of the LDM, indicating an estimated improvement of 83% for I3 and 88% for I4. When compared to the other models mentioned in Table I, the I4 approach outperforms the traditional FRDM and the ML-based MDN by 36% and 16%, respectively. Nevertheless, its global performance is not as strong as the theoretical WS4 model or the ML methods GPR and SVR.

In Fig. 3, we depict the Δ values for both training and testing subsets across the entire nuclide chart to illustrate the local differences. Note that the layout is the same as in Fig. 1, except that the colour bar in the current figure ranges from -3 to 3 MeV, compared to -14 to 14 MeV in the LDM figure. Actually, the change in the colour bar range directly comes from the improvement of nuclear mass predictions. Besides, the predominant colour in both sub-figures is white, meaning that the predicted values are close to the experimental ones. However, the I3 model outputs slightly overestimated masses around the magic numbers, an issue that is partially resolved with the addition of the pairing term in I4. For instance, the isotopic chain with $Z = 50$ has a $\sigma = 0.64$ MeV for I3 and a $\sigma = 0.42$ MeV for I4, indicating an improvement of 34% in this specific chain. Additionally, the I4 sub-figure has a more prominent white tone than the I3, reflecting the superior performance of the former over the latter.

So far, both models have been trained and evaluated on AME2016. Thus, in what follows, we asses their performance on the 120 newly included nuclei in AME2020 to determine their extrapolation competences. The computed outcomes show that I3 achieves a $\sigma = 0.64$ MeV, while I4 outputs $\sigma = 0.57$ MeV, both results being comparable to the ones obtained from the AME2016 testing subsets.

To better illustrate their behaviour in the AME2016 and AME2020 regions, the masses along the isotopic chain of the superheavy element Meitnerium (Mt) are presented in Fig. 4. As shown, both I3 and I4 perform accurately in the two regions, achieving $\sigma = 0.20$ MeV and $\sigma = 0.31$ MeV over the entire chain, respectively. In this case, I3 demonstrates a better performance than I4. In comparison, the LDM and WS4 models yield significantly higher values of $\sigma = 9.26$ MeV and $\sigma = 1.40$ MeV, respectively. These outcomes highlight the strength of the CNN-based models, as they not only surpass traditional models in this particular case but also succeed in maintaining a minimal difference between the experimental values and the predicted ones along the chain, even in the extrapolation region. Notably, while the errors of the LDM and WS4 models systematically increase with the neutron number, both I3 and I4 constantly keep the differences Δ close to zero, emphasizing their robustness and reliability for nuclear mass predictions. Hence, this consistency ensures that future predictions for newly discovered nuclei can be computed with high precision using the CNN architectures.

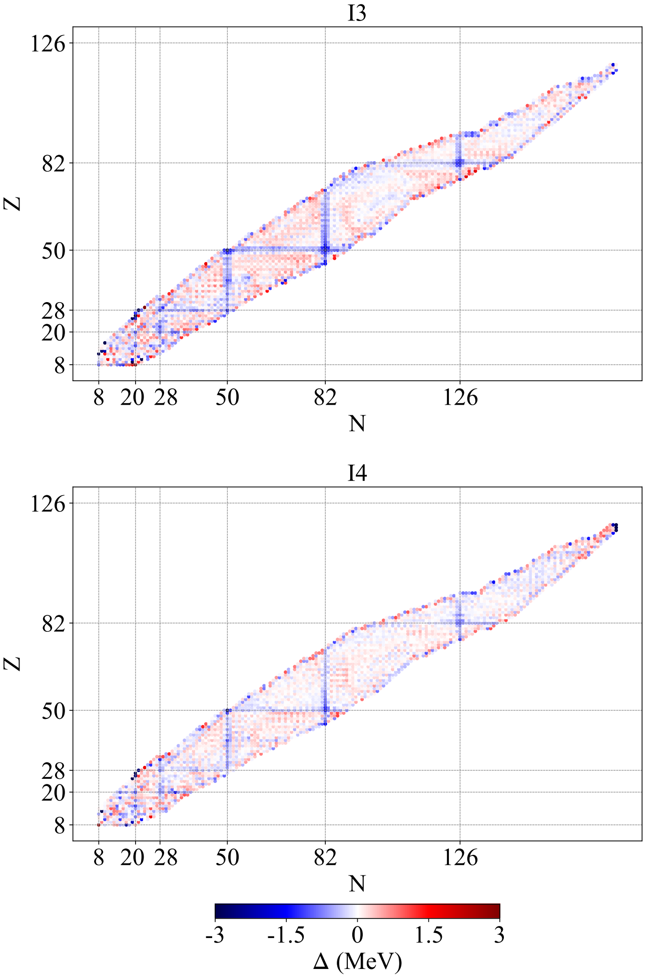


FIG. 3. Scatter plots of Δ (5) between the AME2016 database and I3 (upper) and I4 (lower), including both training and testing subsets.

VII. CONCLUSIONS

In this work, we have demonstrated the successful implementation of CNNs to precisely predict nuclear masses. Our models, I3 and I4, outperform traditional approaches like the LDM by approximately 83% and 88%, respectively. This major improvement emphasizes the ability of CNNs to capture complex patterns within nuclear data that traditional models struggle to identify.

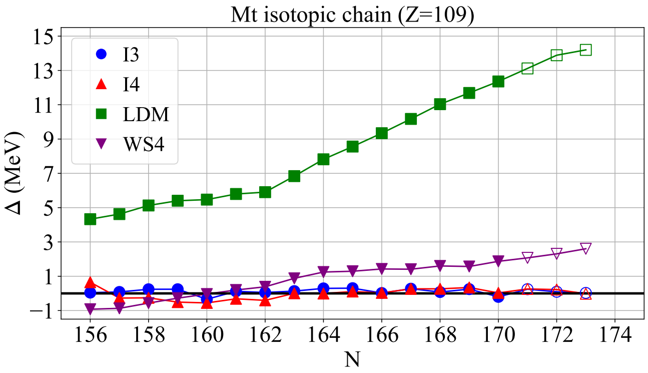


FIG. 4. Isotopic chain of the element Mt ($Z = 109$) and its corresponding values of Δ (5) using different models, each one represented by a different symbol. The filled points depict the AME2016 region, while the empty points indicate the AME2020 extrapolation region.

On the other hand, the I4 model has achieved a 27% of improvement in the σ testing subset with respect to the I3 model. The enhancement is attributed to the introduction of the nuclear pairing effect. Furthermore, the I4 model surpasses the traditional FRDM and the ML-based MDN by 36% and 16%, respectively.

These results highlight the effectiveness of the CNN models, as they are situated in a competitive spot among other approaches. Moreover, they exhibit strong generalization abilities, maintaining high precisions even when extrapolated to newly discovered nuclei in the AME2020 database. This underscores the potential of CNNs as a promising avenue to substantially enhance the precision of nuclear mass predictions.

To conclude, even though the accuracy of 100 keV required for r-process research has not yet been achieved, this work represents a major step forward and establishes the foundations for future developments. As demonstrated, the inclusion of more physical information into the model's input leads to more accurate predictions. Thus, further research should focus on building on this concept. For instance, one promising step could be the creation of a $5 \times 5 \times 5$ tensor as the input for each target nucleus, with the additional layer containing the WS4 theoretical masses. This strategy would allow the CNN-based architecture to benefit from both the local properties of the nuclide chart and the most accurate theoretical predictions to date, ultimately enhancing the precision of nuclear masses.

Finally, the CNN approach demonstrated in this work could also be applied to predict other nuclear observables, such as charge radii or proton and neutron shell gaps, leveraging the spatial structure of the nuclide chart. This could deepen our understanding of nuclear structure and extend predictions to unexplored regions of the chart.

ACKNOWLEDGMENTS

I would like to sincerely thank my supervisor, Dr. Arnau Rios, for his guidance and support throughout this project. I am also grateful to Alejandro Romero for his comments and suggestions. Special thanks go to my sister for her assistance, and to my parents for providing me with this opportunity. Lastly, I would like to express my appreciation to my partner for her emotional support.

[1] M. Mumpower, R. Surman, G. McLaughlin, *et al.*, The impact of individual nuclear properties on r-process nucleosynthesis, Prog. Part. Nucl. Phys. **86**, 86–126 (2016).

[2] D. Guan and J. Pei, High quality microscopic nuclear masses of superheavy nuclei, Phys. Lett. B. **851**, 138578 (2024).

[3] K. Mahboub, A simple theoretical approach to the liquid drop model, Ann. Nucl. Energy **35**, 1381–1385 (2008).

[4] P. Möller, W. Myers, W. Swiatecki, *et al.*, Nuclear mass formula with a finite-range droplet model and a folded-Yukawa single-particle potential, At. Data Nucl. Data Tables **39**, 225–233 (1988).

[5] N. Wang, M. Liu, X. Wu, *et al.*, Surface diffuseness correction in global mass formula, Phys. Lett. B. **734**, 215–219 (2014).

[6] E. Yüksel, D. Soydaner, and H. Bahtiyar, Nuclear mass predictions using machine learning models, Phys. Rev. C **109**, 064322 (2024).

[7] A. E. Lovell, A. T. Mohan, T. M. Sprouse, *et al.*, Nuclear masses learned from a probabilistic neural network, Phys. Rev. C **106**, 014305 (2022).

[8] D. Kumar, Math behind Convolutional Neural Networks (2024), [Online] Accessed: 2024-12-29.

[9] Y. Lu, T. Shang, P. Du, *et al.*, Nuclear mass predictions based on convolutional neural network (2024), arXiv:2404.14948 [nucl-th].

[10] M. Wang, G. Audi, F. Kondev, *et al.*, The AME2016 atomic mass evaluation (II). Tables, graphs and references, Chin. Phys. C **41**, 030003 (2017).

[11] D. Lunney, J. M. Pearson, and C. Thibault, Recent trends in the determination of nuclear masses, Rev. Mod. Phys. **75**, 1021–1082 (2003).

[12] D. Benzaid, S. Bentridi, A. Kerraci, *et al.*, Bethe–Weizsäcker semiempirical mass formula coefficients 2019 update based on AME2016, Nucl. Sci. Tech. **31**, 9 (2020).

[13] K. S. Krane, *Introductory nuclear physics* (Wiley, New York, 1988).

[14] M. Wang, W. Huang, F. Kondev, *et al.*, The AME2020 atomic mass evaluation (II). Tables, graphs and references, Chin. Phys. C **45**, 030003 (2021).

[15] D. Morales, Nuclear mass predictions based on CNN (2025), [Online] Accessed: 2025-01-17.

[16] S. Gureja, Introduction to model optimization in PyTorch (2023), [Online] Accessed: 2025-01-11.