UNIVERSITEIT VAN AMSTERDAM

Bachelor Thesis Physics and Astronomy

Accelerating Lanthanide and Actinide Opacity Calculations with Machine-Learning Models

By:

Khoi Pham

Conducted between March 31st, 2025, and June 19, 2025

Student number:	14591103
Research institute:	ARCNL
Supervisor:	Bowie Brewster & John Sheil
Examiner:	Oscar Versolato

Abstract

Lanthanides and actinides are heavy elements produced in neutron-star mergers that dominate the opacity of kilonova ejecta through complex f-shell line spectra, but calculating their bound–bound opacities with atomic-structure codes is computationally extremely costly. This work introduces a two-stage machine-learning approach that replaces those calculations. To address this, the NIST-LANL opacity database — comprising over 12000 high-resolution spectra spanning 28 elements, 27 temperatures and 17 densities — was collected into a unified dataset for machine-learning. A convolutional autoencoder first reduced a down-sampled 200-bin opacity spectrum to a low-dimensional latent space, which a gradient-boosted regressor then mapped back to opacity. This surrogate was able to reproduce Planck-mean values with errors on the order of 10^3 cm^2/g and reduced spectral RMSE by more than 50% relative to a nearest-neighbour baseline. Next, a convolutional classifier analyzed these spectra, inferring element identity with 99.7% accuracy, plasma temperature within 0.19 eV RMSE, and mass density within 8.35×10^5 RMSE. Although aggressive down-sampling smoothed out the narrowest resonances and performance degraded in the most highly ionized regimes, the surrogate evaluated opacities orders of magnitude faster than conventional atomic-structure codes. Future work on attention-based models to recover the narrow peaks, physics-informed loss functions to enforce atomic constraints, and uncertainty-driven active learning could sharpen accuracy and enable transparent, uncertainty-aware opacity calculations in next-generation astrophysical simulations.

Contents

1 Introduction

1.1 Neutron Star Mergers and Kilonovae

The merger of two neutron stars represents one of the most energetic and scientifically rich phenomena in modern astrophysics. Not only do these mergers emit gravitational waves — first directly detected in 2017 by the LIGO and Virgo collaborations — but they also power a thermal kilonova transient as synthesized, radioactive heavy nuclei in the neutron-rich ejecta decay and heat the expanding material. Kilonovae have since been recognized as the electromagnetic counterparts to gravitational-wave events, confirming that neutron-star mergers are the primary cosmic sites for rapid neutron-capture (r-process) nucleosynthesis, which produces roughly half of the elements heavier than iron—including gold and platinum [1].

Understanding the electromagnetic signatures of these ejecta is crucial for interpreting multi-messenger observations, where gravitational-wave detections are combined with electromagnetic measurements to infer the merger's dynamics and ejecta properties. However, modeling the radiative output of these ejecta remains a challenging task. The opacity of the ejecta is dominated by bound-bound transitions in highly ionized, heavy element ions, giving rise to complex spectra that vary with temperature, density, and composition [2].

1.2 Opacity in Heavy-Element Plasmas

Opacity plays a central role in radiative-transfer models of astrophysical plasmas, determining how light emerging from an explosion or merger is absorbed and scattered before it reaches an observer [3]. The monochromatic opacity, κ_ν, quantifies the probability for photons of frequency ν to be absorbed or scattered per unit mass. In lanthanide- and actinide-rich kilonova ejecta, κ_ν is dominated by four physical processes: bound–bound transitions (the absorption and re-emission of photons as electrons jump between discrete energy levels), bound–free photoionization (photons energetic enough to free an electron into the continuum), free–free bremsstrahlung (radiation produced when free electrons scatter off ions without changing their charge state), and Compton scattering by free electrons.

In lanthanide/actinide-rich ejecta at temperatures $T \sim 2000$–$10000\,\mathrm{K}$ and densities $\rho \sim 10^{-20}$–$10^{-4}\,\mathrm{g\,cm^{-3}}$, the bound–bound component dominates by orders of magnitude [2]. Physically, the open 4f and 5f subshells possess extremely dense manifolds of energy levels. Each ion can have on the order of 10^7 allowed electric dipole transitions; when thermal broadening is included, these transitions blend into a thick forest of lines [4].

1.3 The Planck Mean Opacity

In radiative simulations it is often neither practical nor necessary to carry the full frequency-dependent opacity, $\kappa_\nu(T, \rho)$, at every timestep. For optically thin conditions—where photons escape after at most one interaction—the Planck mean opacity, κ_P, provides a single scalar that weights all photon frequencies by the local thermal emission. It is defined as

$$\kappa_{\mathrm{P}}(T,\rho) = \frac{\displaystyle\int_0^\infty \kappa_\nu(T,\rho) B_\nu(T)\mathrm{d}\nu}{\displaystyle\int_0^\infty B_\nu(T)\mathrm{d}\nu}, \tag{1}$$

where $B_\nu(T)$ is the Planck function. Because $B_\nu(T)$ peaks where the plasma is brightest, κ_{P} effectively weighs the photon energies that contribute most to the thermal emission. In practice, however, the enormous "forest" of bound–bound lines implies that evaluating κ_{P} requires resolving κ_ν at tens of thousands of frequency points.

1.4 The NIST-LANL Opacity database

The National Institute of Standards and Technology (NIST) and Los Alamos National Laboratory have released a comprehensive database of opacity spectra covering 28 lanthanide and actinide elements over 17 densities and 27 temperatures — a total of 12852 individual spectra [5]. Each spectrum is line-binned on a uniform frequency grid spanning the ultraviolet through the near-infrared, with enough resolution to capture the full f-shell line forest of each ion under local thermodynamic equilibrium (LTE).

Due to the complexity of these processes, computing a single bound–bound spectrum at this resolution can require several hours of CPU time on high-performance atomic-structure codes. An accurate surrogate that could reproduce these spectra or their Planck means at an arbitrary input (Z, T, ρ) would be extremely valuable. Such a surrogate could allow radiative-transfer codes to query opacities without needing the multi-hour cost of a new atomic-structure calculation each time.

1.5 Goals of this thesis

The central aim of this work is to replace costly atomic-structure calculations with fast, accurate machine-learning surrogates for bound–bound opacities. Different regression based models are trained on the parameters Z, temperature T, and mass density ρ to accurately output the opacity spectrum. Each model's ability to reproduce sharp the overall continuum will be evaluated using different mean-squared-error metrics. Furthermore, classifiers that ingest a binned spectrum are trained to recover its parent element, temperature bin, or density bin, exploring how well spectral patterns alone encode physical conditions. The ultimate goal of this thesis is to identify which architectures best balance accuracy and computational efficiency, enabling rapid predictions suitable for integration into kilonova radiative-transfer simulations.

2 Machine Learning Models and Methodology

2.1 Introduction to Machine Learning

In this work, machine learning provides a data-driven approach to approximate the complex relationship between plasma conditions and opacity spectra without resorting to expensive atomic-structure calculations at each point. Formally, the task is cast as supervised learning: given a dataset of input–output pairs (x_i, y_i), where each x_i might be a vector of variables (Z, T, ρ) or binned spectrum, and is either the corresponding

full opacity curve or a physical label such as element identity, the aim is to learn a parameterized function $f_\theta(x)$ that minimizes the error

$$\frac{1}{N}\sum_{i=1}^{N}\mathcal{L}(f_\theta(x_i), y_i), \tag{2}$$

where $\mathcal{L}$ is a loss function such as mean-squared error for regression or cross-entropy for classification [6]. The parameters θ are adjusted by iterative optimization methods such as gradient descent until the model works well on held-out data.

Neural networks are a flexible class of models for this task, capable of approximating highly nonlinear mappings through compositions of trainable building blocks called neurons [7]. By arranging neurons into layers one can tailor the network's inductive biases to the structure of opacity data. Architectural choices such as the number of layers, layer widths, convolutional kernel sizes, and activation functions all impact the final accuracy of the model. Throughout this chapter, different neural-network architectures will be explored to establish effective designs for opacity prediction.

2.2 Database and pre-processing

As mentioned earlier the opacity spectra were scraped from the NIST–LANL Lanthanide/Actinide Opacity Database [5]. The public release comprises 12 852 spectra covering 28 lanthanide ($57 \leq Z \leq 71$) and actinide ($89 \leq Z \leq 103$) elements, logarithmically spaced over 27 temperatures ($0.01 - 5$) eV and 17 densities ($10^{-20} - 10^{-4}$) g cm^{-3}. Every spectrum provides the total monochromatic opacity, alongside with the bound–bound, bound-free, free-free, and scattering components, the LTE charge-state distribution and the mean ionic charge. Figure 1 illustrates how the ionization state varies across (T, ρ) for four representative elements. The raw opacity values span many orders of magnitude and exhibit sharp, narrow peaks that can dominate standard loss functions, so two key preprocessing and training strategies were evaluated.

2.2.1 Scaling experiments

Opacity values span many orders of magnitude and feature sharp peaks that can dominate training, so every spectrum was scaled before feeding it into the models. Two transforms were tested: a log transform $\log(\kappa + \epsilon)$, with a small ϵ to avoid $\log(0)$, and a cube root transform, $\kappa^{1/3}$. Previous work has shown that cube-root scaling often reaches a better balance between compressing large peaks and preserving smaller features [8]. In practice, cube-root scaling yielded more stable training, faster convergence, and consistently lower validation errors compared to the log transform, so it was adopted for all subsequent experiments.

2.2.2 Loss metrics

Most machine learning models generally implement the mean squared error loss function

$$\textbf{MSE} = \frac{1}{N}\sum(\hat{x_i} - x_i)^2, \tag{3}$$

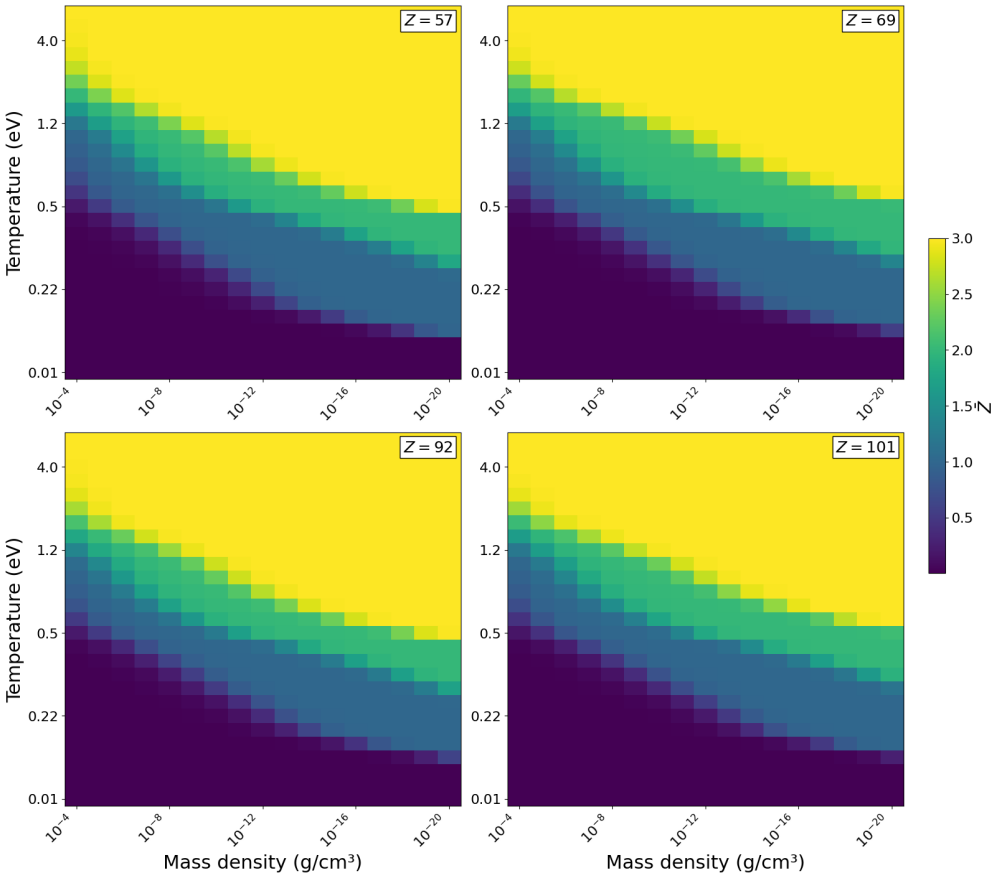

Figure 1: Mean charge state $\bar{Z}$ as a function of temperature and density for four selected elements: $Z = 57$ (La), $Z = 69$ (Tm), $Z = 92$ (U), and $Z = 101$ (Md). Yellow denotes highly ionised plasmas, while dark regions correspond to nearly neutral conditions.

which treats every bin equally. However, in bound–bound opacity spectra the tallest peaks carry the most physical significance and dominate error metrics. To ensure these critical features receive extra emphasis, a gradient-weighted MSE (WMSE) was introduced:

$$\mathrm{WMSE} = \frac{1}{N}\sum_{i=1}^{N} w_i\,(\hat{x}_i - x_i)^2, \quad w_i = 1 + \alpha|x_{i+1} - x_i|. \tag{4}$$

Here, w_i amplifies the loss where the local spectral gradient $|x_{i+1} - x_i|$ is large, which is at the sharpest rises and falls. The hyperparameter α was tuned to balance peak emphasis without destabilizing optimization and fixed at $\alpha = 50$. Models trained with WMSE showed substantially lower errors at the strongest spectral features and improved overall spectral fidelity, so WMSE was adopted alongside standard MSE for all regression experiments.

2.2.3 Binning

Each raw opacity spectrum in the database contains around $N \sim 15000$ energy points ranging from 0.1 eV to 10 eV. This high resolution allows every narrow f-shell transition

to be captured in the spectrum. For the simple regression based networks such as the fully connected neural network the energy was added to the model as an input parameter, meaning that large spectra do not necessarily lead to more complex computing or memory issues. However, for convolutional models the number of operations per layer and the size of stored activations both scale roughly linearly with the input length. In practice, training on the full 15000-point grid would stretch a single-GPU run from minutes per epoch to multiple hours, and force batch sizes well below 32 just to fit in memory.

Therefore, each spectrum was linearly interpolated onto a logarithmic energy grid of only 200 points. This 75× reduction preserves the integrated strength of the broad features that dominate radiative-transfer calculations, while shedding the narrowest spikes. All models except the FC-NN use this 200-point representation.

2.2.4 Training and test data

The full dataset was shuffled and partitioned into a training set (90%) and a validation set (10%), ensuring that both splits uniformly sample the full range of elements, temperatures, and densities. Initial experiments trained separate regression models for each element, which allowed the network to specialize on the temperature–density trends within a single element and showed strong performance. A combined model trained across all elements, however, struggled to understand inter-element patterns, resulting in higher errors. Consequently, all regression results are reported for element-specific models unless otherwise noted. The models developed in this research were implemented using PyTorch 2.0 with fixed random seeds to ensure reproducibility [9].

2.3 Regression models

2.3.1 Nearest-Neighbour Baseline

A straightforward nearest-neighbour baseline was implemented to set a performance floor. For any target temperature–density pair, the four surrounding points on the original NIST–LANL grid are located, and their precomputed opacity values are retrieved. The baseline prediction is simply the mean of these four opacities. This method requires no training, runs in constant time, and returns a mean-squared error that serves as a clear lower bound for any surrogate model.

2.3.2 Fully Connected Neural Network (FC-NN)

First a a fully connected neural network was developed to test the ability to capture nonlinear dependencies between plasma conditions and opacity spectra. In this project, the model took three standardized inputs — energy, temperature and mass density — and returned the corresponding opacity. The architecture consisted of four layers of 512 units, all using ReLU activations [10].

The network was trained to minimize the weighted mean-squared error using the Adam optimizer with an initial learning rate $\eta = 10^{-3}$ and weight decay of 10^{-6} [11]. Data was batched in groups of 8172 samples, over 5 epochs.

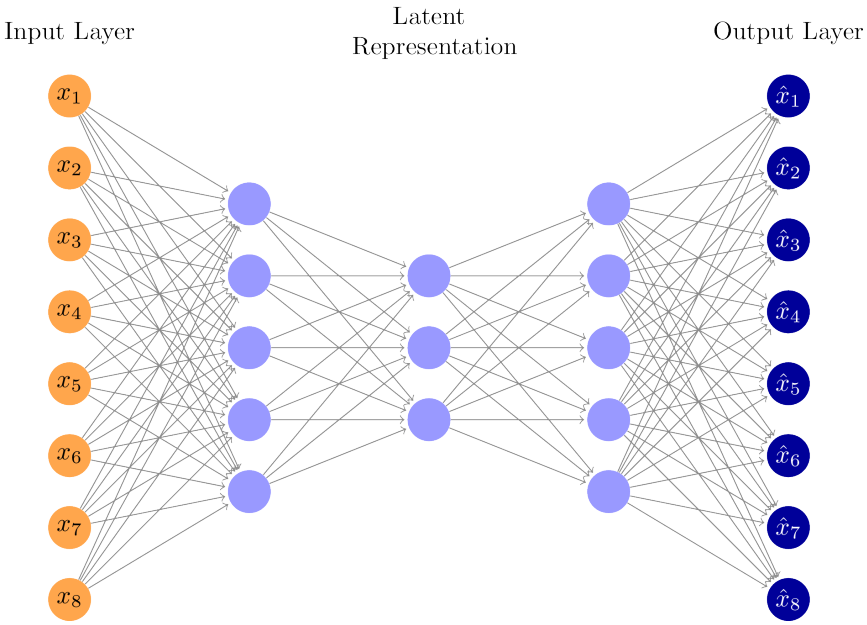

Figure 2: Schematic network representation of the convolutional autoencoder. The encoder (left) compresses the input spectrum to the latent space through successive convolutional layers, while the decoder (right) expands it back to a reconstructed spectrum. Figure adapted from Tikz.net [12].

2.3.3 Convolutional Autoencoder (CAE)

One of the limitations of the point-wise regression with an FC-NN is its inability to capture correlations across an entire opacity spectrum. To address this, applying an autoencoder could allow better pattern recognition within the spectrum [8]. CAEs have multiple convolutional layers to reduce dimensionality, allowing them to capture the key features and patterns within these opacity spectra.

The CAE consists of three parts — an encoder, the latent space, and decoder — as shown in figure 2. Encoders typically consist of multiple layers, decreasing in number of nodes. This forces the network to learn the most important features to store within each layer. The latent space is where the compressed data is stored during training. The number of nodes in this layer can be interpreted as a tunable hyperparameter. Too few nodes could lead to too much loss of data, while too many would fail to capture the most important features. Finally, the decoder reconstructs the original data from the latent space using the key feature stored.

To reduce dimensionality the data was interpolated into a 200-point spectrum as discussed in section 2.2.3. The CAE's encoder then applied three successive 1D convolutional layers (kernel size 5, stride 2, zero padding), growing the channel depths of $32 \to 64 \to 128$, which reduce the spectrum length from $200 \to 100 \to 50 \to 25$. After the final convolution, the resulting 128×5 was flattened passed through a dense layer to a 16-dimensional latent vector. The decoder reversed this process: a linear layer expands the code back to 12825, after which the 200-point spectrum was reconstructed. The CAE was trained on the weighted MSE described using Adam ($\eta = 10^{-3}$), batch size 256, up

to 7 000 epochs.

After training the autoencoder, its weights were fixed and every spectrum was encoded into its 16-dimensional representation. A separate multilayer perceptron — a network with similar structure as the FC-NN — was trained to learn the $(T, \rho) \rightarrow z$ mapping. This Latent-MLP had two hidden layers of 64 ReLU neurons each, followed by a 16-neuron output layer. This network was trained using Adam ($\eta = 10^{-3}$), batch size 64 for 5000 epochs, resulting in $\sim 8,8 \times 10^3$ parameters. The architecture of the model is summarized in table 1

At evaluation, any new temperature-density pair was passed through the Latent-MLP to predict a latent vector z, which subsequently was decoded back to a 200-point spectrum. This architecture allowed the most important spectral patterns to be captured in the CAE to provide a clear mapping from any temperature-density pair to full spectrum.

Table 1: Compact architecture summary for the CAE and Latent-MLP.

CAE	
Input dim.	200
Latent dim.	16
Encoder conv. layers	$3 \times (k = 5, s = 2) : 32 \rightarrow 64 \rightarrow 128$
Decoder conv-tp layers	$3 \times (k = 5, s = 2) : 128 \rightarrow 64 \rightarrow 32$
CAE params	$\approx 1.1 \times 10^5$
Latent-MLP	
Input features	2 (T, ρ)
Hidden layers	2×64 (ReLU)
Output layer	16 (linear)
MLP params	$\approx 8.8 \times 10^3$

2.3.4 LightGBM Regression

Although the CAE and MLP approach delivers a direct mapping from plasma conditions to opacity spectra, it introduces practical challenges. The MLP that predicts latent codes from temperature and density often demands careful tuning of learning rates, weight decay, batch size and early-stopping criteria; they can be sensitive to initialization and prone to overfitting. Training also takes longer than for tree-based methods, and the resulting network behaves like a "black box," making it hard to tell how temperature or density individually shape each latent dimension.

To overcome the tuning sensitivity and interpretability challenges of neural regressors, LightGBM—a gradient-boosting framework developed by Microsoft—was employed to map temperature and density directly to the CAE's latent codes [13]. LightGBM grows trees leaf-wise, selecting splits that maximize loss reduction, and uses histogram-based feature binning along with optimizations such as Gradient-based One-Side Sampling and Exclusive Feature Bundling to accelerate training and mitigate overfitting. Its tree-based structure also yields built-in feature-importance metrics, allowing direct quantification of how each input influences individual latent dimensions [14].

After training the autoencoder, each temperature-density pair was fed through an LGBM-regressor to predict a latent vector z, which was then passed through the decoder to recover a 200-point spectrum. Without extensive hyperparameter search, the regression objective was configured to 500 trees, a learning rate of 0.05, and up to 31 leaves per tree. The LGBM architecture is summarized in table2.

Table 2: Hyperparameter and setup summary for the LightGBM-based latent regressor.

LightGBM regression	
Input features	2 (temperature T, density ρ)
Target	16-dimensional CAE latent code
Wrapper	`MultiOutputRegressor`
Objective	regression
`n_estimators`	500
`learning_rate`	0.05
`num_leaves`	31
Early stopping	100 rounds on validation loss

2.3.5 Planck Mean Opacity Evaluation

Because many radiative–transfer solvers require only a single scalar opacity per zone, the performance was not only measured by the root mean-squared error (RMSE) of the entire spectrum, but also on ability to reproduce the Planck mean opacity κ_p. As mentioned in the introduction the Planck mean κ_p weights each monochromatic opacity κ_ν by the Planck function, emphasizing the thermally most important peaks. For each test case, the models predicted values κ_ν were substituted into equation 1, which was then compared with the true Planck mean.

2.4 Classification models

Inferring physical conditions and element compositions from opacity spectra both validates the encoder's latent representations and enables inverse modeling applications. If a single trained encoder can reliably predict the parent element, temperature, and density from an opacity spectrum, it has effectively captured the key spectral signatures of each ion and plasma state. Such classifier could identify dominant species and conditions in observed kilonova spectra and guide active sampling strategies in atomic-structure calculations.

A multitask convolutional network was used to perform all three tasks in one framework. The network takes a spectrum and encodes it through four sequential convolutional layers. Similar to section 2.3.3, these layers expand the opacity spectrum to a 128×5 feature map. This feature map then is flattened into a 640-dimensional vector and projected through a connected ReLU to form a 128-dimensional latent vector. From this latent vector three branches are created, training the element, temperature and mass density simultaneously.

2.4.1 Element Classification

The element classification head maps the 128-dimensional latent vector to 28 nodes — one for each lanthanide and actinide — and applies a softmax to produce a probability distribution over atomic numbers. During training, categorical cross-entropy loss is weighted inversely to each element's frequency in the training set to correct for class imbalance. The model is optimized using the Adam algorithm with an initial learning rate of $\eta = 10^{-3}$), which is halved whenever validation accuracy plateaus. Training was done in 128 batches and early stopping is triggered once validation accuracy shows no improvement for 50 epochs.

2.4.2 Temperature and density classification

The temperature and density branch each consists of a single linear projection from the 128-dimensional latent vector. One branch outputs the temperature in eV, the other one outputs the exponent of the density $\log_{10} \rho$. Both branches share the same parameters as the classifier branch. By sharing the encoder weights and training all three branches jointly, the model was trained to learn a unified representation that captures the spectral features needed for classification and regression alike. The architecture of the model is summarized in table 3.

Table 3: Multitask Convolutional-NN Architecture Summary

Shared Encoder	
Input	200-point cube-root–scaled spectrum
Conv layers	4×(kernel=5, ReLU): 1→16→32→64→128; strides: 2,2,2,5
Flatten & Dense	128×5 → 640 → 128 (ReLU) latent space
Element Branch	
Output	Dense(128→28) + softmax
Loss	Weighted categorical cross-entropy
Temperature Branch	
Output	Dense(128→1) linear (eV)
Loss	Mean-squared error on T
Density Branch	
Output	Dense(128→1) linear ($\log_{10} \rho$)
Loss	Mean-squared error on $\log_{10} \rho$
Training Setup	
Optimizer	Adam ($\eta = 1 \times 10^{-3}$, reduce on plateau)
Batch size	128
Early stopping	50 epochs patience on validation

3 Results

3.1 Baseline vs. Neural Nets vs. LightGBM (Regression)

3.1.1 FC-NN Performance

The fully connected neural network provided a straightforward non-linear mapping from (E, T, ρ) to the opacity value κ_ν. To demonstrate it's behaviour on a single test case, figure 3 shows four different spectra with the predictions on Praseodymium (Z = 59). Even though the FC-NN is roughly able to capture the general trend of the spectra, it is clear that the model fails to understand the line-forests that are prevalent in the bound-bound opacity, resulting in an average error of RMSE $= 2.8 \times 10^7$ cm^2/g. This error is too large for any application requiring precise line-by-line opacity. The core limitations of this approach stem from its lack of any inductive bias toward locality in frequency space and its purely global, dense mapping of inputs to outputs, which together prevent the model from resolving the sharp, closely spaced transitions that form the line forest. Given these findings, it is clear that purely dense, feed-forward networks are inadequate for modeling bound-bound opacities.

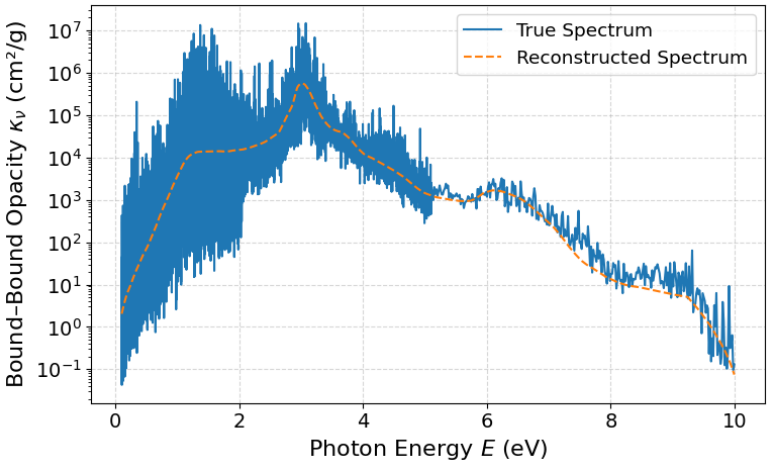

Figure 3: True vs. predicted spectrum of Praseodymium (Z = 59) for a random temperature-density pair ($T = 0.17$ eV, $\rho = 10^{-14}$ g/cm^3) using a fully connected neural network.

3.1.2 Autoencoder Performance

As detailed in section 2.3.3, a convolutional autoencoder was trained to learn a compact, frequency-localized representation of the bound–bound opacity spectra, followed by a small MLP that maps (T, ρ) into the CAE's latent space before decoding back to the full 200-point spectrum. It is important to note that by binning each high-resolution line forest down to just 200 energy bins, much of the finest forest structure is already smoothed out before any modeling, meaning the CAE is working on an input that has lost its narrowest resonances.

When evaluated on similar Praseodymium test conditions, the CAE and MLP model nevertheless is able to quite accurately reconstruct a the binned Praesodymium spectrum as shown in figure 4. Especially in the higher opacity regions — which are the most relevant — the predicted lines are very similar to the true lines. In addition, for Praesodymium the model was able to predict the Planck mean opacity κ_p with an error of RMSE $= 2.6 \times 10^3$ cm^2/g. Over all the lanthanides and actinides, the model was able to predict the opacity with an average error of RMSE $= 2.2 \times 10^5$ cm^2/g, nearly a factor of two improvement over the nearest neighbour baseline of RMSE $= 4.1 \times 10^5$ cm^2/g. This improvement demonstrates how the convolutional filters of the autoencoder recover much of spectral patterns, making the CAE + MLP a substantially more accurate surrogate than both simple nearest-neighbour averaging and fully connected neural networks.

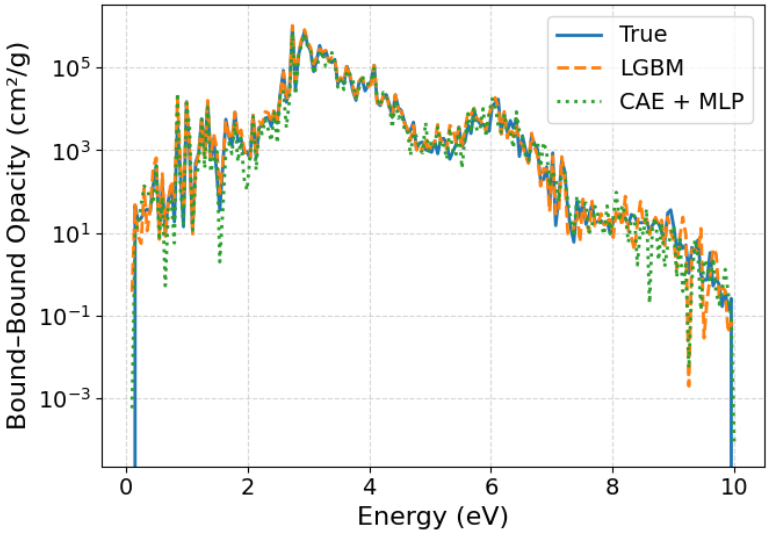

Figure 4: True vs. predicted spectrum of Praseodymium (Z = 59) for a random temperature-density pair ($T = 0.27$ eV, $\rho = 10^{-15}$ g/cm^3) using an CAE with both MLP and LGBM.

3.1.3 LightGBM Performance

Building on the CAE framework from Section 2.3.3, the Latent-MLP was replaced by a LightGBM multi-output regressor to map (T, ρ) into the 16-dimensional latent space before decoding back to the full 200-point spectrum. On the Praseodymium test case, this CAE+LightGBM pipeline achieves an average RMSE of 1.4×10^3 cm^2/g on the Planck mean, which is significantly lower than the standard MLP model. Overall this model achieves an average accuracy of RMSE = 1.6 $\times10^5$ cm^2/g and decreased the Planck mean error by $\sim 20\%$. As also seen in figure 4, the LightGBM-based decoder matches or slightly outperforms the MLP version in capturing the surviving troughs of the binned line forest, highlighting the benefits of a tree-based latent regressors for convolutional representation.

To determine which input feature—temperature or density—contributes most to this gain, feature-importance scores were averaged across all LGBM regressors and plotted versus atomic number in figure 5. Temperature importance consistently lies near 43×10^3 A.U., while density importance remains near 3 $\times10^3$ A.U. throughout the lanthanides and actinides.

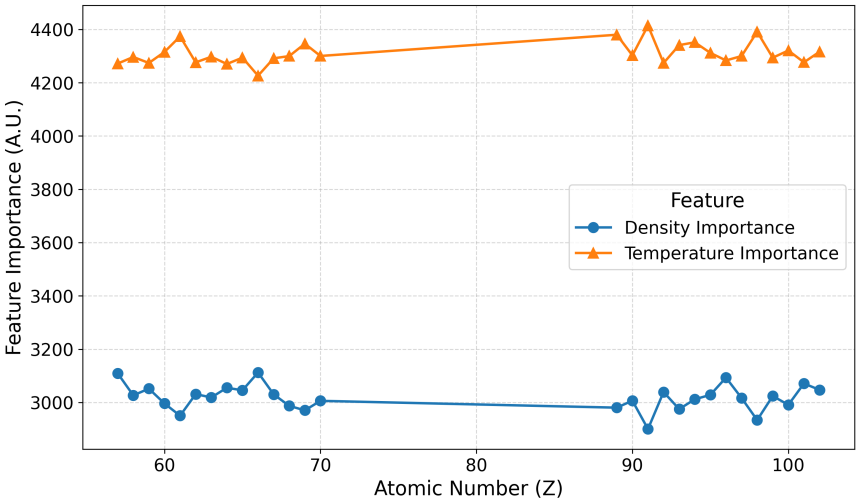

Figure 5: Average LightGBM feature-importance for predicting each dimension of the CAE latent code, plotted against atomic number. Temperature consistently contributes more to loss reduction than density across all 28 lanthanide and actinide elements, highlighting temperature's dominant influence on the opacity spectra.

3.1.4 Overall comparison

Finally, figure 6 brings together the nearest-neighbour baseline, CAE + MLP, and CAE + LGBM by plotting validation RMSE versus atomic number. The nearest-neighbour approach exhibits large, element-specific error spikes. Adding a CAE significantly improves the models predictions, while LightGBM delivers the lowest and most uniform errors across the most of the elements. This hierarchy shows that encoding frequency locality via convolutional autoencoding — and then refining the mapping from state variables to latent space with gradient-boosted trees — produces the most accurate surrogates for bound–bound opacities of the lanthanides and actinides.

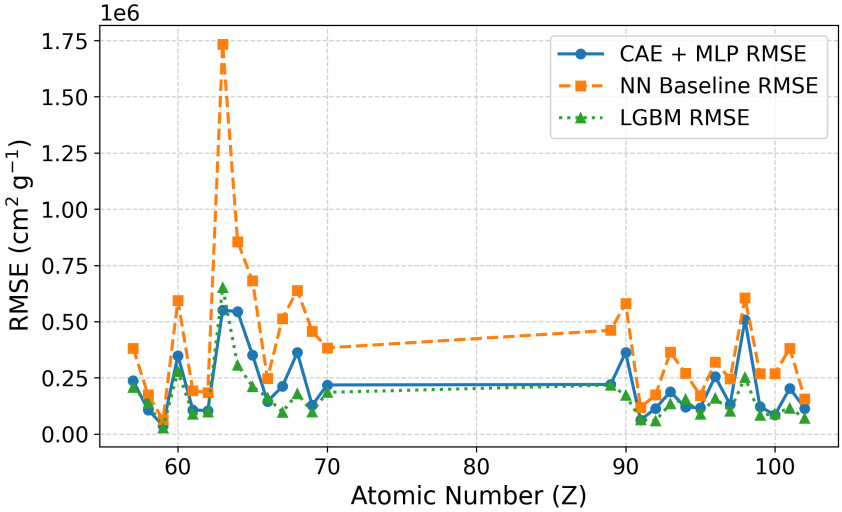

Figure 6: Validation RMSE vs. atomic number. The NN baseline shows large, element-specific spikes in error; both CAE + MLP and CAE + LightGBM deliver substantially lower and more uniform performance across all 28 lanthanide and actinide elements.

3.2 Classification Results

The multitask convolutional network excelled at extracting physically meaningful features from the 200-point binned spectra, delivering accurate element identification fine temperature estimates.

On the held-out test set, the element-classification branch achieved an accuracy of 99.7%, demonstrating that the network was able to recognize the distinct line-forest patterns of each lanthanide and actinide.

The temperature branch recovers the plasma temperature with an RMSE of 0.19 eV, as shown in the left plot in figure 7. The final density branch was able to recover the density with an RMSE of 8.35×10^{-5} g cm^{-3} accuracy, corresponding to an RMSE of 2.57 on the exponent. This is a slightly worse performance compared to temperature, which also can be seen in the right plot in figure 7. Especially in the higher and lower extremes of the density the scatter grows, while deviating from the true values. These results correspond to the findings from the LGBM regression in section 3.1.3, which indicated that temperature carried a higher feature importance than density. The final performance metrics are summarized in table 4.

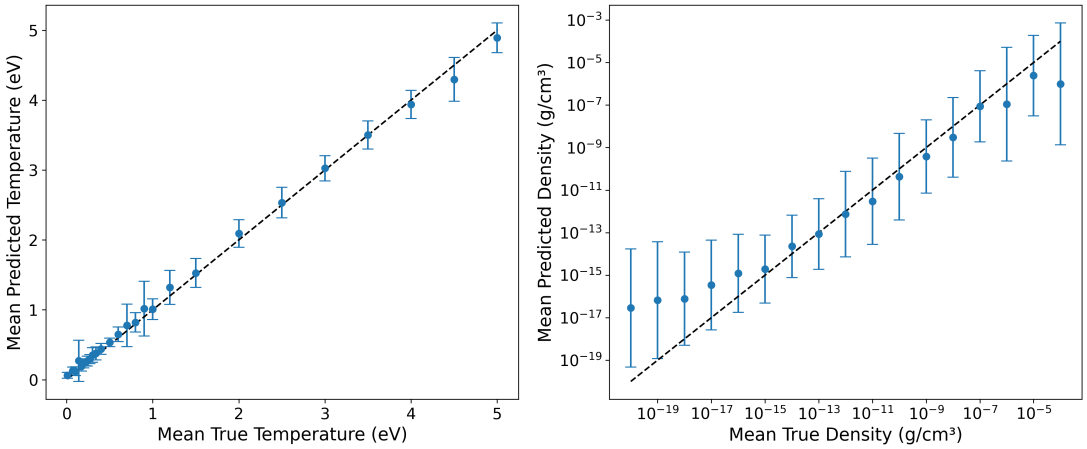

Figure 7: Left: Mean predicted vs. mean true temperature (eV) with error bar. Right: Mean predicted vs. mean true density (g/cm^3); the dashed line marks perfect equality.

Table 4: Validation metrics for the multitask classifier.

Metric	**Value**
Element ID Accuracy	0.997
Temperature RMSE	0.19 eV
Density RMSE	8.35×10^{-5} g/cm^3 (2.57 dex)

4 Discussion

4.1 How Well Do ML Models Capture Spectral Features?

The fully connected neural network (FC-NN) was trained on the unbinned, full-resolution spectra of $\sim$ 15000 points, hoping to understand every narrow f-shell interaction. In practice however, those sharp, densely packed line forests overwhelmed the dense architecture. The FC-NN frequently misaligned individual peaks and misestimated their heights, essentially performing as a bad fit function. The average RMSE of 2.8×10^7 cm^2/g is in the order of the peaks of the opacities, meaning on average the model's prediction error was as large as the very features it was intended to reproduce. This level of discrepancy makes clear that a feed-forward network, without any mechanism for recognizing local patterns, struggles to model the sharp and dense structure of bound–bound opacity spectra.

By contrast, the convolutional autoencoder pipelines were trained on spectra interpolated down to 200 bins, smoothing the dense 15000-point forest into broader features. This blurring sacrificed the narrowest spikes but retained the overall features that dominate the total opacity. Even though the CAE never sees the original ultrafine lines, its convolutional filters learn the general trend of the line forest and reconstruct peak positions to within a resolution of $\frac{\Delta E}{E} \approx 10^{-2}$. With a LightGBM latent regressor and

gradient-weighted training, it achieved a spectral RMSE of $\sim 1.6 \times 10^5$ cm^2 and reproduced Planck-mean opacity errors of $\sim 10^4$. Practically, this translates to a fractional error of around $\sim 1\%$, well within the tolerances of current kilonova radiative-transfer models.

In summary, the convolutional-autoencoder combined with a LightGBM latent regressor provided a highly efficient and accurate surrogate for bound–bound opacities in kilonova ejecta. The model excelled at recognizing peak locations, but due to the binning it consistently underestimated the strength of those peaks. The resulting surrogate generally achieved a Planck-mean fractional error of approximately 1%, well below the intrinsic uncertainties of current kilonova light-curve and spectral models. Consequently, this approach enables opacity evaluations that are several orders of magnitude faster than direct atomic-structure calculations without introducing significant biases into astrophysical predictions.

4.2 Binning & Information Loss

Down-sampling from ∼15000 to 200 bins inevitably smooths away most of the finest f-shell resonances: most of the individual lines merge into larger bins or disappear entirely. Consequently, the ML surrogates never directly see these narrow peaks, leading to systematic underestimation of their heights.

This aggressive compression is driven by practical considerations. In the convolutional-autoencoder architecture, both compute effort and memory usage scale approximately linearly with the input length. Working at full resolution would increase training time from hours to days or weeks and force batch sizes so small that optimization stability would suffer. By down-sampling to 200 points, epoch runtimes were cut by more than an order of magnitude, sustain batch sizes of several hundred, and keep GPU memory well within the limits of our hardware (an 11th Gen Intel Core i7-1165G7 CPU with Intel Iris Xe GPU). On more powerful systems, intermediate resolutions of 1000–2000 bins could be explored to recover additional spectral detail without prohibitive cost. However, attempting to process the full 15000-point spectra remains impractical even on high-end clusters, both because of extreme memory demands and the very long training times.

Crucially, although fine-scale spikes disappear, the results have shown that models trained on the 200-bin spectra were able to retain more than 97% of the variance relevant to Planck-weighted opacity calculations, meaning that the impact on astrophysical models is minimal, as Planck-mean errors remained around 1%.

For applications that require full line-by-line resolutions, a practical compromise could be to restore the missing sharp peaks on a range basis rather than calculating entire spectra. The on-demand calculation option could use the surrogate's own error estimate, meaning if its predicted uncertainty for a given temperature-density point crosses a certain tolerance, an original atomic-structure code could be triggered for that single location. This hybrid approach could still significantly reduce computational cost, while still delivering accurate predictions.

4.3 Classification Takeaways

The multitask convolutional network has demonstrated exceptional capability in extracting and leveraging physically meaningful features from binned opacity spectra. In partic-

ular, the element-classification branch achieved an accuracy of 99.7% on the held-out test set, confirming that the network reliably recognizes the unique signatures of each lanthanide and actinide. Such near-perfect performance indicates that the encoder's latent representations are highly discriminative with respect to atomic number.

For continuous-valued targets, the network recovers plasma temperature with an RMSE of 0.19 eV, reflecting high precision across the bulk of the test set. However, the standard error on the temperature predictions increases noticeably around $\sim$ 4 eV, a regime in which the mean ionization state $\bar{Z}$ rises sharply as seen in figure 1. Physically, as the plasma becomes more highly ionized, many of the lower-energy bound–bound transitions vanish and the remaining lines shift toward fewer, more widely spaced high-charge-state features, while the bound-free interactions strengthen [4, 15]. With fewer diagnostic lines and a stronger bound-free background, the network has less spectral information to constrain fine-scale temperature variations, which could produce the increase in temperature residuals observed at higher-$\bar{Z}$ regions.

Density estimation yielded an RMSE of 8.35×10^{-5} g cm^{-3}, with performance degrading at the extremes of the density range. This behavior mirrors the feature-importance analysis from the LGBM regression (section 3.1.3), which attributed greater predictive weight to temperature-sensitive spectral features than to density-sensitive ones.

The success of a single encoder to perform element classification, temperature regression, and density regression demonstrates its potential for inverse modeling of kilonova ejecta. Specifically, such classifier could potentially identify the dominant heavy-elements in observed spectra, infer local plasma conditions from these spectra, and guide targeted opacity calculations by highlighting regions of parameter space where uncertainty remains high. Nevertheless, the uneven performance in temperature and density estimation still highlights some limitations. While convolutional feature extractors excel at capturing sharp line-forest patterns, they may require architectural or loss-function adjustments to fully understand variations associated with temperature and density. In the next subsection different extensions for future work will be explored to target these issues.

4.4 Future Extensions

4.4.1 Attention-Based Neural Networks

One of the more straightforward improvements that could be made to the models explored in this research is combining convolutional encoders with self-attention to further improve the model's performance in high-$\bar{Z}$ regimes. Convolutional encoders perform well with understanding local structures, but struggle when features become sparse and separated, which happens in those high-$\bar{Z}$ regimes. Self-attention layers allow each segment dynamically to weigh information from the entire input, creating a global view over the entire spectrum while retaining the local pattern-recognition strengths of convolutions [16].

Because attention need only operate over a small number of segments, the additional parameters and compute cost remain relatively low, yet could significantly increase performance on high-ionization spectra. Similar convolution-attention mechanisms have already been proven effective in other spectral and imaging applications, such as in UV-Vis spectrum prediction, working-memory spectroscopy, and infrared spectral forecasting [17, 18, 19].

In the context of kilonova opacity, implementing self-attention could both reduce the current model's tendency to underestimate the narrow peaks, and improve the temperature and density estimates in precisely the high-$\bar{Z}$ regimes where the current models show the largest residuals.

4.4.2 Physics-Informed Neural Networks (PINNs)

Physics-Informed Neural Networks (PINNs) allow ways to apply known physical relations directly into the training of the model. Instead of solely relying on supervised losses, PINNs add extra penalty when deviated from known physical equations [20].

First, a simple loss that enforces that as temperature increases, the strength and density of bound-bound line features systematically decrease [4]. This could be done by defining a metric within a spectrum as

$$B_k = \frac{1}{|P|} \sum_{j \in P} \kappa_{k,j}, \tag{5}$$

where P indexes the top $p\%$ opacity bins with the strongest line features. Because the bound-bound opacity decreases with increasing temperature, when sorting a training batch any increase in B_k should be penalized:

$$\mathcal{L}_{bb} = \frac{1}{B-1} \sum_{m=1}^{B-1} \max(0, B_{(m+1)} - B_m), \tag{6}$$

where $B_{(m)}$ is the metric at the mth-lowest temperature in the training batch.

In addition, an extra branch could be trained to predict the mean charge state $\bar{Z}$. With the Saha ionization equation,

$$\frac{N_{i+1}\, n_e}{N_i} = \frac{2}{\lambda_{\text{th}}^3} \frac{g_{i+1}}{g_i} \exp\left[-\frac{\chi_{i+1} - \chi_i}{k_B\, T} \right], \tag{7}$$

the mean charge $\bar{Z}_{\text{Saha}}$ can be determined from temperature and density, with a relatively cheap computational cost [21]. By including a mean-squared error loss on $\bar{Z} - \bar{Z}_{\text{Saha}}$, the encoder could learn features that respect ionization balance, increasing accuracy on both temperature and density predictions.

Combining these loss functions with the standard supervised loss function, the combined loss function becomes

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \gamma \mathcal{L}_{\text{bb}} + \delta \mathcal{L}_{\bar{Z}}, \tag{8}$$

where γ and δ weight the respective loss influence. By enforcing the known decline of bound-bound line strength with temperature and the known Saha relation, a PINN-enhanced network is guided to learn real physical features instead of blindly relying on the black-box nature of standard encoders.

4.4.3 Active Learning

An active-learning framework could further accelerate both the expansion opacity database and the improvement of new ML-model performance by identifying regions where its own

uncertainty is highest and then requesting full atomic-structure calculations only at those critical (T, ρ) points. For example, the convolutional encoder's latent uncertainty could be estimated by LightGBM regressors and flag regions of parameter space where predictions are least reliable. Once new opacities are generated by atomic-structure calculations at those locations they could be added back to the training set, iteratively improving the model exactly where it needs it most.

Such targeted sampling avoids wasting computational effort on well-understood regimes while closing the gaps in the current models. In chemometric regression, active learning has been shown to reduce the number of labeled spectra required for accurate prediction by iteratively querying the most informative samples [22]. Similarly, log-Gaussian process–driven active learning has been used to optimize neutron spectroscopy experiments by focusing on conditions with the highest uncertainty [23].

5 Conclusions

To assess whether machine-learning surrogates can understand and reproduce lanthanide and actinide opacities, we collected the entire NIST-LANL opacity database into a comprehensive training set spanning all relevant elements, temperatures and densities. Different regression models were constructed, working with autoencoders and gradient-boosted regressors, to reconstruct spectra. In addition, a convolutional classifier was built to infer elements, temperature, and density based on a single opacity spectrum. Throughout this process, different parameters, loss functions, and architectures were compared to identify the optimal approach.

The convolutional autoencoder paired with the LightGBM regressor was generally able to predict the Planck-mean opacity κ_p with an error in the order of $\sim 10^3$ cm^2/g, within a few percent of general Planck-mean opacities, and achieved an average RMSE of 1.6 $\times 10^5$ cm^2/g, more than a twofold improvement over the nearest-neighbour baseline. A multitask convolutional classifier then used the same spectra to infer element identity with 99.7% accuracy, temperature within 0.19 eV RMSE, and mass density to within 8.35×10^{-5} g/cm^3 RMSE. Together these results confirm that the CAE + LGBM architecture and classifier companion can serve as a faithful, fast alternative for full atomic-structure codes, reproducing spectral shapes and mean opacities within general modeling tolerances.

Despite these strong results two key limitations emerged. Aggressive down-sampling from ∼15000 to 200 bins removed the narrowest f-shell resonances and led to systematic underestimation of peak opacities in the most line-dense regions. The classifier, while highly accurate overall, showed increased temperature residuals in the ∼4 eV — where the ionization state is higher and bound-free processes dominate — and greater scatter in density predictions in general. To address these limitations, future work could integrate self-attention networks over full-resolution wavelength sequences, enabling the model to focus adaptively on narrow line clusters; incorporate physics-informed neural network (PINN) loss terms that embed known atomic physics constraints to sharpen line predictions; and embed calibrated uncertainty estimation to drive an active-learning loop that requests new full-physics calculations only where model confidence is lowest.

In closing, this framework demonstrates a clear path for replacing costly atomic computations with machine-learning surrogates that balance speed and accuracy, while

also offering new insights into the underlying physics. By highlighting which spectral features drive opacity and quantifying model uncertainty, it not only accelerates opacity calculations but deepens our understanding of the bound–bound processes. This lays the groundwork for responsive, uncertainty-aware opacity calculations and more transparent physical interpretation in next-generation astrophysical studies.

References

[1] Friedrich-Karl Thielemann et al. "Neutron Star Mergers and Nucleosynthesis of Heavy Elements". In: *Annual Review of Nuclear and Particle Science* 67 (2017), pp. 253–274. DOI: 10.1146/annurev-nucl-101916-123246. URL: https://doi.org/10.1146/annurev-nucl-101916-123246.

[2] C J Fontes et al. "A line-binned treatment of opacities for the spectra and light curves from neutron star mergers". en. In: *Monthly Notices of the Royal Astronomical Society* 493.3 (Apr. 2020), pp. 4143–4171. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/staa485. URL: https://academic.oup.com/mnras/article/493/3/4143/5758310 (visited on 06/03/2025).

[3] George B. Rybicki and Alan P. Lightman. *Radiative Processes in Astrophysics*. New York: Wiley, 1979. ISBN: 978-0471848695.

[4] C. J. Fontes et al. "Actinide opacities for modelling the spectra and light curves of kilonovae". In: *MNRAS* 519 (2023), pp. 2862–2878.

[5] Yu. Ralchenko et al. *NIST-LANL Lanthanide Opacity Database (ver. 1.2)*. [Online]. [Accessed May 12, 2025]. Gaithersburg, MD 20899, 2025. DOI: 10.18434/mds2-2375. URL: https://nlte.nist.gov/OPAC.

[6] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[7] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. Available online at http://www.deeplearningbook.org. MIT Press, 2016.

[8] Michael D. Vander Wal, Ryan G. McClarren, and Kelli D. Humbird. "Neural network surrogate models for absorptivity and emissivity spectra of multiple elements". In: *Machine Learning with Applications* 8 (June 2022), p. 100308. DOI: 10.1016/j.mlwa.2022.100308.

[9] PyTorch Contributors. *PyTorch*. https://pytorch.org/. Accessed: 2025-06-19.

[10] Kaiming He et al. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification". In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 2015, pp. 1026–1034.

[11] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. arXiv preprint arXiv:1412.6980. 2014.

[12] Izaak Neutelinks. *Neural networks*. June 2025. URL: https://tikz.net/neural_networks/.

[13] Guolin Ke et al. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree". In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon et al. Curran Associates, Inc., 2017, pp. 3146–3154. URL: http://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf.

[14] *Features - LightGBM 4.6.0.99 documentation*. URL: https://lightgbm.readthedocs.io/en/latest/Features.html#optimization-in-speed-and-memory-usage.

[15] Sultana N. Nahar et al. "Highly Excited Core Resonances in Photoionization of Fe XVII: Implications for Plasma Opacities". In: *arXiv preprint arXiv:1104.2881* (2011). URL: https://arxiv.org/abs/1104.2881.

[16] Ashish Vaswani et al. "Attention Is All You Need". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*. arXiv preprint arXiv:1706.03762. 2017, pp. 6000–6010. URL: https://arxiv.org/abs/1706.03762.

[17] Fabio Urbina et al. "UV-adVISor: Attention-Based Recurrent Neural Networks to Predict UV–Vis Spectra". In: *Analytical Chemistry* 93.48 (2021), pp. 16076–16085. DOI: 10.1021/acs.analchem.1c03741. URL: https://doi.org/10.1021/acs.analchem.1c03741.

[18] Xiang Guo et al. "Programming ability prediction: Applying an attention-based convolutional neural network to functional near-infrared spectroscopy analyses of working memory". In: *Frontiers in Neuroscience* 16 (2022), p. 1058609. DOI: 10.3389/fnins.2022.1058609. URL: https://www.frontiersin.org/articles/10.3389/fnins.2022.1058609/full.

[19] Naseem Saquer et al. "Infrared spectra prediction using attention-based graph neural networks". In: *Digital Discovery* 3 (2024), pp. 602–609. DOI: 10.1039/D3DD00254C.

[20] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations". In: *Journal of Computational Physics* 378 (2019), pp. 686–707. DOI: 10.1016/j.jcp.2018.10.045. URL: https://doi.org/10.1016/j.jcp.2018.10.045.

[21] Francis F. Chen. *Introduction to Plasma Physics and Controlled Fusion*. 3rd ed. Cham, Switzerland: Springer International Publishing, 2016. ISBN: 978-3-319-22308-7. DOI: 10.1007/978-3-319-22309-4.

[22] Fouzi Douak et al. "Active Learning for Spectroscopic Data Regression". In: *Journal of Chemometrics* 26.7 (2012), pp. 381–389. DOI: 10.1002/cem.2443.

[23] Mario Teixeira Parente et al. "Active Learning-Assisted Neutron Spectroscopy with Log-Gaussian Processes". In: *Nature Communications* 14 (2023), p. 2246. DOI: 10.1038/s41467-023-37418-8. URL: https://www.nature.com/articles/s41467-023-37418-8.