

# Recoverability of Atomic Quantum Structure from Emission Spectra

Cameron Gattis

Independent Researcher, Jarrell, TX, USA  
gattiscm@gmail.com (corresponding author)  
cameron.gattis@g.austincc.edu  
ORCID: 0009-0007-5950-9468

## Abstract

Atomic emission spectra encode physical structure through observable photon energies, yet the inverse mapping from spectra to atomic configuration is inherently information-entropic. In this work, we empirically assess which aspects of atomic quantum structure are recoverable from spectral observables alone and which are irretrievably compressed under radiative projection by quantifying information retention and loss. Using the NIST Atomic Spectra Database [1], we first demonstrate near-perfect statistical recovery of the Planck–Einstein relation, establishing a baseline for invertible physical structure. We then show that global reconstruction of Moseley–Rydberg scaling fails decisively for neutral multi-electron atoms, reflecting genuine degeneracy rather than model inadequacy. Constrained neural network models are subsequently used as diagnostic probes of information content, revealing a clear hierarchy of recoverability: initial-state quantum numbers exhibit statistically significant recoverability in restricted atomic regimes, while final-state quantum numbers are generally unstable and non-invertible, collapsing toward chance under mixed-element conditions. These results demonstrate that radiative emission preserves only partial and regime-dependent information about atomic structure, with systematic loss increasing under electronic screening and configuration mixing.

# 1 Introduction and Background

Atomic emission spectra provide one of the most direct experimental windows into microscopic structure, encoding information about electronic energy differences through observable photon frequencies and wavelengths. However, the mapping from atomic configuration to emitted radiation is inherently information-entropic. Photon emission represents a projection from a high-dimensional electronic state space onto a one-dimensional observable, and the invertibility of this projection depends sensitively on the underlying physical regime.

At the most fundamental level, the Planck–Einstein relation establishes a linear and invertible mapping between photon frequency and energy [2, 3],

$$E = h\nu, \tag{1}$$

independent of atomic structure. This relation holds universally and does not depend on the internal composition of the emitting system. As a result, photon energy can be uniquely recovered from spectral observables alone, making the inverse problem well posed.

In contrast, atomic structure enters spectroscopy through conditional relations that depend on discrete quantum numbers and nuclear charge. In hydrogenic systems, the Rydberg formula expresses photon energy as a function of the principal quantum numbers of the initial and final states [4, 5]. Moseley’s law generalizes this scaling to higher nuclear charge by incorporating an approximate  $Z^2$  dependence [6], reflecting Coulomb attraction between the nucleus and a single effective electron. These relations are exact or nearly exact only in systems where electron–electron interactions are negligible or can be treated perturbatively.

For neutral multi-electron atoms, however, screening, configuration mixing, relativistic effects, and angular-momentum-dependent splittings introduce substantial degeneracy [5, 7]. Distinct electronic configurations can emit photons with nearly identical energies, and the same photon energy can arise from multiple transition pathways across different elements. In such regimes, photon energy information alone no longer uniquely identifies the underlying atomic state. The inverse problem becomes ill posed: information about the initial configuration is partially preserved, while information about the final configuration is often erased under radiative projection.

This distinction between invertible and non-invertible spectral mappings motivates a reframing of spectroscopic reconstruction. Rather than asking whether a single analytic law applies globally, one may ask which physical parameters leave statistically detectable imprints on spectral observables and which are fundamentally compressed. From this perspective, the failure to reconstruct a law is not necessarily a deficiency of the model, but evidence of genuine information loss. This interpretation is strengthened by explicit tests across alternative model activations and classifiers, which preserve the observed hierarchy of recoverability while failing to restore invertibility for final-state quantities.

While the qualitative non-invertibility of atomic emission spectra in multi-electron systems is well established, the present work addresses a distinct and previously unquantified question: which components of atomic quantum structure retain measurable mutual information with spectral observables under strictly spectral-only constraints, and which are irretrievably lost under radiative projection. To our knowledge, no prior study has empirically mapped this hierarchy of recoverability across atomic species while explicitly forbidding access to atomic identifiers, quantum numbers, selection rules, or imposed analytic structure. The contribution of this work is therefore not the discovery of a new spectroscopic law, but the quantitative delineation of information retention and loss in atomic emission as a function of electronic complexity.

Machine-learning methods provide a useful diagnostic tool in this context. When constrained to spectral observables alone, such models cannot invent physical structure; they can only exploit information present in the data. Their successes and failures therefore serve as empirical probes of the information content encoded in emission spectra. In this work, machine learning is used not as a replacement for analytic physics, but as an instrument for testing recoverability, degeneracy, and the limits of inversion across atomic systems [8].

All analyses are presented cumulatively. Earlier models are not replaced or removed when extended; instead, each subsequent step is appended to expose both the persistence and the breakdown of underlying assumptions.

## 2 Dataset and Dataset Preparation

All analyses in this work are based on photon emission data obtained from the NIST Atomic Spectra Database, using the publicly available CSV exports [1]. Two dataset scopes were maintained throughout: a global dataset aggregating all available elements after cleaning (`cleaned_elements.csv`), and element-specific subsets corresponding to individual atomic species (e.g., H, He, C, Ne, Na, Ca, Ti, Fe, Hg, Pb). Raw CSV files were ingested into `pandas` DataFrames [9], with standardized column naming and UTF-8 encoding cleanup applied uniformly to remove byte-order marks and formatting inconsistencies.

Prior to any modeling, the dataset was subjected to systematic cleaning and sanitization to remove non-physical or ambiguous entries. Rows with missing or non-numeric wavelength, frequency, or energy values were discarded, as were transitions lacking defined lower or upper state annotations. All spectral observables were explicitly coerced to numeric form, including wavelength, frequency, and photon energy in electron volts. Duplicate spectral lines arising from redundant NIST listings of identical transitions were removed to prevent artificial weighting. These steps ensure that subsequent analyses operate exclusively on physically meaningful photon transitions.

To further suppress spurious structure arising from poorly resolved measurements, uncertainty-based filtering was applied. NIST uncertainty annotations were parsed, and transitions exceeding a predefined uncertainty threshold were excluded. Only spectral lines with reliable wavelength or frequency precision and stable transition labeling were retained. This filtering step is essential for preventing models from learning artifacts associated with measurement noise rather than atomic structure.

Derived physical quantities were appended to the dataset without overwriting original NIST fields. Photon frequency was computed from wavelength according to

$$\nu = \frac{c}{\lambda},$$

and photon energy was computed using the Planck–Einstein relation

$$E = h\nu.$$

For analyses involving mass–energy correspondence, the equivalent mass change was calculated as

$$\Delta m = \frac{E}{c^2}.$$

All derived quantities were stored as additional columns, preserving the original observables to ensure traceability and reproducibility.

Atomic number  $Z$  was incorporated through a separate isotope lookup table, enforcing a one-to-one mapping between element symbols and nuclear charge. Isotope-specific identifiers were deliberately excluded unless explicitly required for a given analysis, in order to avoid unintended overfitting to nuclear mass variations. The inclusion of  $Z$  enables Moseley-type scaling tests,  $Z^2$  normalization experiments, and controlled cross-element aggregation.

Quantum state information provided by NIST was encoded non-destructively. Rather than replacing original spectroscopic annotations, additional columns were introduced for the principal quantum numbers  $n_i$  and  $n_k$ , and for the orbital angular momentum quantum numbers  $\ell_i$  and  $\ell_k$ . Orbital labels were mapped as  $s \rightarrow 0$ ,  $p \rightarrow 1$ ,  $d \rightarrow 2$ ,  $f \rightarrow 3$ , and  $g \rightarrow 4$ , while the original string annotations were retained unchanged. This design choice preserves backward compatibility with earlier models, enables parallel neural network experiments, and facilitates independent verification by reviewers.

For selected evaluation runs, physics-aware masking based on electric dipole selection rules was applied [5, 10]. Specifically, transitions violating

$$\Delta\ell = \pm 1$$

were excluded during evaluation but never removed from the underlying dataset. This separation prevents data leakage while allowing physics-constrained diagnostics and transparent assessment of failure modes.

A secondary preprocessing pass generated density-based auxiliary features, including local photon density in energy or frequency space, neighbor counts within adaptive bins, and transition clustering indicators. These features were written to a separate directory and used only as auxiliary inputs in selected models, never as replacements for primary spectral observables.

To diagnose where reconstruction succeeds or fails across atomic regimes, element-wise dataset partitions were constructed. Hydrogen serves as a baseline one-electron control, helium as the first many-body system, light elements such as carbon, neon, and sodium as weakly screened systems, transition metals such as calcium, titanium, and iron as strongly correlated regimes, and heavy atoms such as mercury and lead as relativistic, high-degeneracy cases. This partitioning allows physical breakdowns to be distinguished from modeling artifacts.

Immediately prior to modeling, feature scaling was applied exclusively to input variables using standard normalization procedures, while labels were left unscaled. Train-test splits were performed only after all filtering and feature construction steps were completed, ensuring that no information leakage or artificial correlations were introduced during preprocessing.

## 3 Methodology

The methodological design of this study was guided by a single principle: all inference must be performed using spectral observables alone. Atomic identifiers, quantum numbers, selection rules, and analytic relationships were explicitly excluded from model inputs. This constraint ensures that any recovered structure reflects information genuinely encoded in emission spectra rather than imposed physical priors.

### 3.1 Dataset Preparation

All analyses were based on CSV exports from the NIST Atomic Spectra Database. A global dataset containing all cleaned transitions was constructed alongside element-specific subsets for targeted

diagnostics. Raw tables were ingested into structured data frames, with column names standardized and encoding artifacts removed. Only transitions with physically meaningful wavelength, frequency, and energy entries were retained.

Rows containing undefined upper or lower states, non-numeric values, or duplicate spectral lines were removed. Measurement uncertainty annotations provided by NIST were parsed explicitly, and transitions exceeding a fixed uncertainty threshold were excluded to prevent poorly resolved lines from biasing learning. These steps ensure that the dataset represents stable, physically interpretable photon emissions.

## 3.2 Feature Construction and Augmentation

Derived quantities were added as new columns without overwriting original observables. Photon frequency was computed from wavelength, photon energy from frequency using the Planck relation, and a mass-equivalent energy scale was included for auxiliary analyses. Atomic number was merged from external reference tables using element symbols, with care taken to avoid isotope-specific assumptions unless explicitly stated.

Quantum state labels provided by NIST were preserved in their original form and supplemented by non-destructive numerical encodings of principal and orbital quantum numbers. These encodings were used exclusively as training targets or evaluation labels. At no point were quantum numbers, atomic number, or selection rules included as model inputs.

## 3.3 Physics-Aware Masking and Density Features

For selected evaluations, electric dipole selection rules were applied as masks during analysis rather than as hard constraints during training. This separation prevents information leakage while enabling physics-consistent diagnostics of failure modes.

A secondary preprocessing pass generated density-based features describing local clustering in energy and frequency space. These features quantify spectral crowding and degeneracy and were supplied as auxiliary inputs in selected models. Raw spectral observables were always retained and never replaced by derived features.

## 3.4 Modeling Constraints and Evaluation

All models were trained on standardized input features, with scaling applied only to inputs and never to targets. Train-test splits were performed after all filtering and feature construction steps to prevent leakage. Evaluation metrics were chosen to emphasize structural recoverability rather than predictive optimization.

Neural networks were deliberately constrained in depth and capacity to reduce memorization and encourage learning of global structure [8]. Multiple activation functions and loss formulations were evaluated to test architectural sensitivity; results were invariant under these choices. Model performance was interpreted diagnostically: high accuracy indicates recoverable physical information, while systematic failure indicates genuine non-identifiability. Under these constraints, model performance may be interpreted as an empirical proxy for the mutual information between spectral observables and atomic structure. Persistent failure across architectures is therefore taken as evidence of physical non-identifiability rather than insufficient model expressivity.

Under these constraints, any recovered relationships or quantum-state predictions arise solely from information present in spectral observables. Conversely, persistent failure across architectures

and losses is treated as evidence of intrinsic information loss under radiative projection rather than model inadequacy.

## 4 Goals

The objective of this work was not to introduce a new spectroscopic law or predictive model, but to examine the extent to which physical structure can be recovered from spectral observables alone. All models developed herein were restricted to spectral quantities as inputs, with no access to atomic identifiers, quantum numbers, or imposed physical relationships. Within this constraint, the investigation reduces to three primary goals:

- statistical reconstruction of the Planck-Einstein relation from spectral data,
- attempted reconstruction of Moseley’s law under the same spectral-only constraint, and
- identification of which physical parameters are empirically recoverable from spectral data alone.

All models were trained using the full NIST Atomic Spectra Database in order to establish statistically stable baselines across elements and transitions. While the dataset provides associated physical identifiers and quantum numbers, these quantities were used exclusively as training targets or evaluation labels. At no point were atomic number, quantum numbers, or selection rules provided as model inputs; all inference was performed using spectral observables alone.

The first goal was to test whether the Planck-Einstein relation could be recovered statistically without imposing its functional form. Specifically, the model sought to reconstruct the relationship

$$E = hf = \frac{hc}{\lambda}, \quad (2)$$

using photon frequency or wavelength as the sole input. Successful recovery of this relation serves as a baseline validation that fundamental physical structure can emerge from spectral data alone.

The second goal was to attempt a corresponding reconstruction of Moseley’s law, expressed in proportional form as

$$E \propto Z^2 \left( \frac{1}{n_k^2} - \frac{1}{n_i^2} \right), \quad (3)$$

again without providing atomic number, quantum numbers, or selection rules as inputs. Unlike the Planck–Einstein relation, this law depends implicitly on multi-electron structure, screening effects, and quantum degeneracies, and is therefore not expected to be fully recoverable from spectral energy alone. Partial or failed reconstruction in this case is treated as an informative result rather than a model deficiency.

The final goal was to generalize this inquiry using neural network models, explicitly probing which physical parameters are recoverable from spectral observables and which are fundamentally degenerate. This goal reframes model performance not as predictive success or failure, but as an empirical test of the information content encoded in spectral data.

## 5 Model 1: Artificial Planck–Einstein Relation Reconstruction

As a baseline test of invertible physical structure, the first model was designed to recover the Planck–Einstein relation using spectral observables alone. Because the target relation is linear and one-dimensional, only two features were retained. The dependent variable was constructed as the net change in atomic mass–energy associated with each transition, computed from the difference between the tabulated initial and final atomic energy levels,

$$\Delta E = E_i - E_k.$$

This quantity represents the internal mass–energy change of the atom during emission. Since the emitted photon energy must be non-negative, the target photon energy was defined as

$$E_\gamma = |E_i - E_k|.$$

The sole input feature was wavelength. Two wavelength representations provided by NIST were considered independently: the observed air wavelength, `obs_wl_air`, and the Ritz wavelength, `ritz_wl_air`. The observed wavelength corresponds to direct experimental measurement, whereas the Ritz wavelength is derived from level energies and may therefore introduce artificial consistency. Separate regressions were performed for each case to assess the impact of this distinction. After all preprocessing and filtering steps, the resulting dataset contained 117 638 transitions with two retained columns.

Linear regression of photon energy against frequency yields near-perfect reconstruction of the Planck–Einstein relation in both wavelength representations. The fitted slopes, intercepts, and coefficients of determination are summarized in Table 1.

Target	Slope (eV s)	Intercept (eV)	$R^2$
Ritz	$4.1343517 \times 10^{-15}$	$8.41 \times 10^{-5}$	0.99999998
Observed	$4.1343475 \times 10^{-15}$	$8.80 \times 10^{-5}$	0.99999999
use_wl	$4.1356684 \times 10^{-15}$	$-5.954 \times 10^{-7}$	0.99999999

Table 1: Linear regression results for Planck–Einstein reconstruction using Ritz and observed wavelengths.

Using CODATA-assigned values [11], Planck’s constant expressed in electron-volt seconds is

$$h = 4.135667696 \times 10^{-15} \text{ eV} \cdot \text{s}.$$

The empirically recovered slopes agree with this value to within approximately 0.03% for both wavelength representations. The fitted relation may therefore be written as

$$E = hf + b, \tag{4}$$

where the intercept  $b \approx 8.6 \times 10^{-5}$  eV represents a small systematic offset attributable to experimental uncertainties and spectroscopic conventions rather than a deviation from Planck’s law.

Figure 1 shows the reconstructed energy–frequency relationship across the full dataset. The near-unity coefficient of determination confirms that, when the underlying physics is linear and invertible,

the inverse mapping from spectral observables to physical law is well posed and recoverable without imposing functional form.

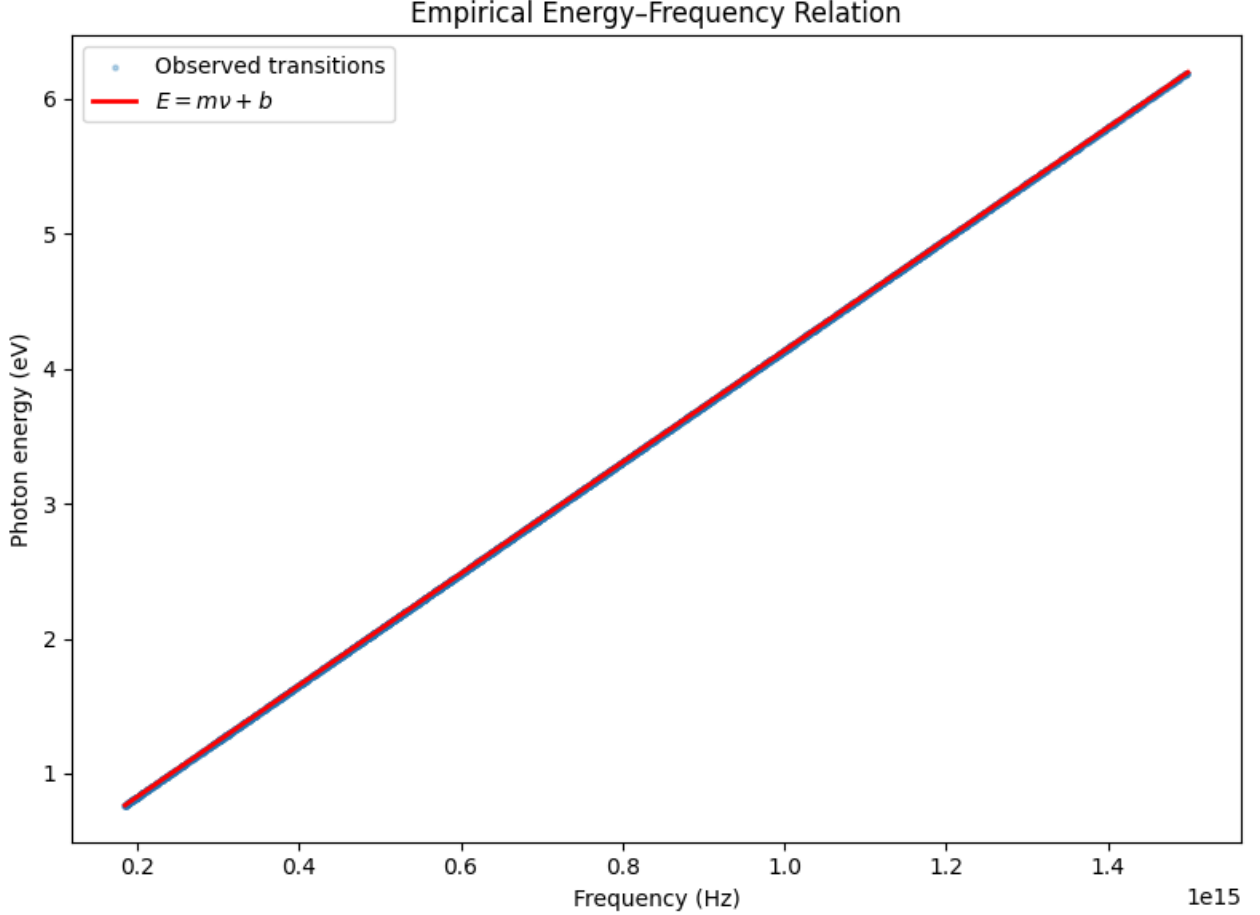


Figure 1: Empirical reconstruction of the Planck–Einstein relation from spectral observables.

## 6 Model 2: Artificial Moseley Law Reconstruction

Following the successful recovery of the Planck–Einstein relation, a second model was constructed to test whether Moseley-type scaling could be reconstructed under the same spectral-only constraint [6]. In contrast to the linear and globally invertible Planck relation, Moseley’s law encodes conditional structure arising from nuclear charge, electronic configuration, and screening effects, and therefore provides a stringent test of inverse recoverability.

As a preliminary check, a naïve linear regression of photon energy against atomic number was performed. This yields a slope of  $1.46 \times 10^{-3}$ , an intercept of 3.13 eV, and a coefficient of determination  $R^2 = 1.37 \times 10^{-3}$ , demonstrating essentially no explanatory power. This failure is expected, as Moseley scaling is not linear in  $Z$  itself, but depends on  $Z^2$  modulated by discrete quantum structure.

Motivated by this, a maximally general linear reformulation consistent with hydrogenic scaling was tested,

$$E = \beta_k \frac{Z^2}{n_k^2} + \beta_i \frac{Z^2}{n_i^2} + \epsilon, \quad (5)$$



where  $n_i$  and  $n_k$  denote the initial and final principal quantum numbers, respectively, and  $\epsilon$  is a constant offset. This form explicitly incorporates all assumptions intrinsic to Moseley’s original derivation: nuclear charge dependence, principal quantum numbers, and linear superposition of initial and final state contributions. Fitting this model yields coefficients of order  $10^{-3}$  and an overall explanatory power of  $R^2 = 7.1 \times 10^{-3}$ . No weaker linear assumptions remain to be relaxed, yet the reconstruction remains ineffective.

The failure of global reconstruction arises from strong isoenergetic degeneracy: distinct atomic configurations project onto nearly identical photon energies. As a result, Moseley-type scaling cannot serve as a global predictor of absolute photon energy without additional discrete state information. When photon energies are rescaled by  $Z^2$ , a residual linear dependence on

$$\Delta\left(\frac{1}{n^2}\right)$$

emerges across mixed atomic transitions, yielding a slope of approximately 1.95 eV and  $R^2 \approx 0.15$ . The reduced slope relative to the hydrogenic Rydberg value reflects strong screening and many-electron effects. Thus, Moseley scaling survives globally only as a weak statistical tendency rather than as a predictive law.

This behavior contrasts sharply with the Planck–Einstein relation,

$$E = h\nu,$$

which is linear in a directly measured observable and therefore globally invertible. In that case, photon frequency uniquely determines energy. Moseley–Rydberg scaling, by contrast, expresses photon energy through a multiplicative power-law dependence on structural quantities,

$$E \propto Z^2 \Delta\left(\frac{1}{n^2}\right),$$

rendering the inverse problem intrinsically many-to-one once electronic structure is present.

Restricting the reconstruction to individual elements further clarifies the domain of validity. Hydrogen reproduces the Rydberg relation with near-perfect fidelity, yielding a slope of 13.6 eV and  $R^2 \approx 1$ . Neutral helium provides insufficient data and is not hydrogenic, rendering the procedure inapplicable. Lithium exhibits a slope near the hydrogenic value but with low explanatory power, reflecting partial screening and configuration mixing. Strongly multi-electron atoms such as iron show no recoverable single-element law. These results demonstrate that reconstructing a universal Moseley or Rydberg relation is only well posed for effectively one-electron systems [7, 12]; once electron–electron interactions dominate, photon energy ceases to uniquely encode atomic structure.

Element	Regime	$N$	Slope (eV)	$R^2$
H	Hydrogenic	89	13.60	0.9999998
Li	Weakly screened	143	14.60	0.046
Fe	Multi-electron	3546	32.06	0.022

Table 2: Element-wise regression results for Rydberg–Moseley scaling tests.

Table 6 illustrates the progressive breakdown of Rydberg–Moseley scaling with increasing electronic complexity. Hydrogen reproduces the Rydberg slope with near-perfect fidelity, confirming exact invertibility in the one-electron Coulomb problem. Lithium exhibits a slope comparable to

the hydrogenic value but with sharply reduced explanatory power, reflecting partial screening and configuration mixing. In iron, both slope and fit quality lose physical meaning entirely, indicating that no single-element analytic scaling survives in strongly correlated, multi-electron systems. The collapse of  $R^2$  with increasing atomic complexity demonstrates that the failure of global Moseley reconstruction is not statistical but structural: photon energy ceases to uniquely encode atomic configuration once electron–electron interactions dominate.

These results demonstrate that the failure of global Moseley reconstruction is not methodological but physical. Once electronic screening and configuration mixing are present, the inverse mapping from photon energy to atomic structure becomes ill posed.

## 7 Model 3: Machine-Learning Assessment of Quantum-State Recoverability

The failure of global analytic reconstruction in the preceding section does not arise from an unknown functional form, but from non-identifiability induced by latent atomic structure. Photon energy represents a many-to-one projection of electronic configuration, and once electron–electron interactions are present, multiple distinct quantum states can map to nearly identical spectral signatures. Recovering conditional physical laws therefore requires inferring which discrete state variables remain encoded in the spectrum after radiative projection. This motivates the use of machine learning not as a substitute for physical theory, but as an empirical diagnostic for identifying which aspects of atomic structure are recoverable from spectral observables alone.

All machine-learning models were restricted to photon-level inputs derived from wavelength and associated measurement metadata. Atomic number, quantum numbers, and selection rules were never provided as inputs and were used only as evaluation labels. Within this formulation, model performance directly reflects the information content of the spectrum rather than algorithmic expressivity. Across all tested activations, loss functions, and classifier formulations, the same ordering of recoverability was observed indicating that observed limitations arise from physical degeneracy rather than model choice.

The principal quantum number of the initial state,  $n_i$ , exhibits substantial recoverability. Using only Ritz wavelength and measurement uncertainty as inputs, models achieve approximately 74% exact-class accuracy across the six most populated principal quantum numbers. All reported accuracies are accompanied by balanced accuracy to account for strong class imbalance; exact-class accuracy alone is therefore interpreted diagnostically rather than as a measure of predictive performance. This performance is far above chance and persists across elements and modeling variants. Lower principal quantum numbers are recovered with the highest fidelity, while higher levels collapse into increasingly compressed spectral manifolds. This behavior reflects the well-known decrease in level spacing with increasing  $n$  and indicates that the models are learning genuine spectral structure rather than memorizing labels.

In contrast, the final-state principal quantum number  $n_k$  is not recoverable. Accuracy collapses toward chance even when electric dipole selection rules are enforced during evaluation. Increasing atomic number, introducing auxiliary density-based features, or modifying loss functions does not restore invertibility. This asymmetry between initial and final states reflects the directional loss of information under emission: while the initial configuration imprints strongly on the emitted photon, the final configuration is largely erased by the many-to-one nature of radiative transitions.

Orbital angular momentum exhibits intermediate behavior. In hydrogen, angular momentum quantum numbers are weakly recoverable, consistent with exact Coulomb degeneracy. In helium and

heavier elements, partial recoverability of the initial-state angular momentum  $\ell_i$  emerges, reflecting angular-momentum-dependent energy shifts induced by electron-electron interaction. Nevertheless, recovery remains incomplete and saturates well below deterministic inversion. The corresponding final-state quantum number  $\ell_k$  shows weaker recoverability still, mirroring the behavior observed for  $n_k$ .

Across all models, confusion matrices display a pronounced columnar structure, indicating the collapse of multiple physically distinct quantum states into observationally equivalent spectral signatures. Alternative loss functions designed to address class imbalance shift macro-level metrics but do not alter this qualitative behavior. These results demonstrate that inverse recovery of final-state quantum numbers from emission spectra is physically underdetermined, even under enforced selection rules.

Target	Classes	Exact Accuracy	Chance	Interpretation
$n_i$	6	$\sim 0.74$	$\sim 0.17$	Strong initial-state imprint
$n_k$	6	$\sim 0.18$	$\sim 0.17$	Not recoverable
$\ell_i$	5	$\sim 0.35$	$\sim 0.20$	Partial recoverability
$\ell_k$	5	$\sim 0.22$	$\sim 0.20$	Largely degenerate

Table 3: Recoverability of quantum-state variables from spectral observables alone. Chance accuracy reflects uniform random guessing over populated classes; reported accuracies correspond to exact-class recovery. Absolute accuracy values are reported for diagnostic comparison only. The primary result is the relative hierarchy of recoverability across quantum variables and atomic regimes, which remains stable across model variants

Several elements, including nitrogen and oxygen, collapse entirely after physically consistent filtering, leaving no statistically meaningful training or evaluation population. These cases are treated as negative controls rather than modeling failures. Their collapse confirms that, for certain atomic regimes, no inverse problem is well defined under spectral-only constraints. Partial recoverability of angular momentum is confined to low-lying initial states; final-state angular momentum remains non-invertible and collapses to effective manifolds.

Taken together, these findings establish a clear hierarchy of quantum-state recoverability. Initial-state structure leaves a detectable imprint on emitted spectra, while final-state structure is largely lost under radiative projection. Recoverability degrades systematically with increasing electronic complexity and spectral degeneracy, indicating genuine information compression rather than model inadequacy. Neural network models recover precisely this hierarchy without access to atomic structure input, providing an empirical confirmation of long-established principles in atomic physics.

## 8 Recoverability of Atomic Quantum Structure from Spectral Observables

The objective of the machine-learning analysis was not to construct a predictive surrogate for atomic structure, but to empirically test which components of that structure are recoverable from spectral observables alone. All models were therefore restricted to photon-level quantities derived directly from the NIST Atomic Spectra Database, with no access to atomic identifiers, quantum numbers, selection rules, or imposed physical relationships. Quantum numbers were used exclusively as diagnostic labels during evaluation. Within this formulation, model performance directly reflects the information content of the spectrum rather than algorithmic capacity.

$Z$	Element	$n_i$ (bal.)	$n_k$ (bal.)	$\ell_i$ (bal.)	$\ell_k$ (bal.)	Regime
1	H	1.00	0.25	$\sim 0.55$	$\sim 0.33$	Hydrogenic (invertible)
3	Li	$\sim 0.32$	$\sim 0.28$	$\sim 0.06$	$\sim 0.30$	Weak screening
6	C	$\sim 0.31$	$\sim 0.29$	$\sim 0.09$	$\sim 0.32$	Partial recovery
10	Ne	$\sim 0.27$	$\sim 0.47$	$\sim 0.22$	$\sim 0.43$	Degenerate
26	Fe	$\sim 0.32$	$\sim 0.66$	$\sim 0.22$	$\sim 0.55$	Multi-electron
49	In	$\sim 0.31$	$\sim 0.03$	$\sim 0.46$	$\sim 0.44$	Initial-state survives
56	Ba	$\sim 0.30$	$\sim 0.25$	$\sim 0.34$	$\sim 0.15$	Sparse manifold
80	Hg	—	—	—	—	Collapsed / undefined

Table 4: Representative balanced accuracy for quantum-state recoverability from spectral observables alone. Chance levels are  $\approx 0.17$  for  $n$  (six classes) and  $\approx 0.20$  for  $\ell$  (five classes). Dashes indicate regimes where physically consistent filtering yields no well-defined inverse problem.

Across all elements examined, a strong asymmetry emerges between initial- and final-state recoverability. The principal quantum number of the initial state,  $n_i$ , is consistently recoverable at accuracies far exceeding chance. In hydrogen, recovery is exact, reflecting the unique invertibility of the one-electron Coulomb problem. As atomic number increases, recovery degrades systematically but remains well above chance, with models converging toward the dominant low-lying shells that contribute most strongly to observed emission. This behavior persists across all modeling regimes, including unconstrained, combined, and canonical variants, indicating that the result is robust to architectural choice.

In contrast, the final-state principal quantum number  $n_k$  is not recoverable in any element examined. Accuracy and balanced accuracy collapse toward chance levels even when electric dipole selection rules are enforced during evaluation. Increasing atomic number does not improve performance, nor does the introduction of density-based auxiliary features. This failure is not a limitation of model expressivity but a consequence of physical degeneracy: multiple distinct final states project onto nearly identical photon energies, rendering the inverse mapping fundamentally ill-posed. The systematic nature of this failure across all tested elements confirms that final-state inversion from spectral observables alone is not physically well defined.

Orbital angular momentum exhibits intermediate behavior. The initial-state angular momentum quantum number  $l_i$  is partially recoverable, with performance improving from hydrogen to heavier elements as electron–electron interactions lift exact Coulomb degeneracy. Nevertheless, recovery remains incomplete, and performance saturates well below deterministic inversion. The corresponding final-state quantum number  $l_k$  shows weaker recoverability still, consistent with the same many-to-one projection that suppresses  $n_k$  inference. The asymmetry between  $l_i$  and  $l_k$  mirrors that observed for the principal quantum number and reflects the directional loss of information under radiative emission.

Application of selection-rule masking improves physical consistency but does not materially alter recoverability. Enforcing  $\Delta\ell = \pm 1$  during evaluation reduces the available test population but does not restore invertibility for final-state quantities. This confirms that the observed failures are not artifacts of forbidden transitions contaminating the dataset, but intrinsic consequences of spectral compression. Canonical model variants yield marginal stability improvements but do not alter qualitative outcomes, further reinforcing that information bottlenecks, rather than modeling choices, govern the results.

Taken together, these findings demonstrate a clear hierarchy of recoverability. Initial-state quantum structure leaves a detectable imprint on emitted spectra, while final-state structure is

largely erased by the radiative projection itself. This apparent loss of information does not imply any violation of unitarity or fundamental information conservation: the combined atom–field system evolves reversibly, but tracing over unobserved atomic and field degrees of freedom renders the photon spectrum a reduced, non-invertible marginal in which mutual information with the initial configuration is necessarily diminished. Machine-learning models recover precisely this hierarchy without explicit physical guidance, converging toward low-dimensional attractors corresponding to dominant atomic manifolds. The resulting behavior is therefore not a statistical curiosity but an empirical reflection of atomic structure and degeneracy.

## 9 Results and Discussion

Taken together, the three modeling regimes establish a consistent picture of which physical structure is recoverable from atomic emission spectra and which is irretrievably lost under radiative projection. When the underlying physical relation is linear and invertible, as in the Planck–Einstein relation, reconstruction from spectral observables succeeds with near-perfect fidelity. Photon frequency uniquely determines energy, and no hidden state variables are required to invert the mapping. This serves as a baseline demonstration that spectral data alone can encode and reveal fundamental physical laws when the inverse problem is well posed. That this reconstruction succeeds is not itself surprising; its role is to establish that the methodology recovers physical structure precisely when the underlying mapping is invertible. These failures are invariant under model architecture and therefore reflect physical non-identifiability rather than methodological limitation.

Moseley-type scaling, by contrast, illustrates the breakdown of global invertibility once atomic structure becomes conditionally encoded. Although hydrogenic systems recover the Rydberg relation exactly, attempts to reconstruct a universal Moseley law across neutral atoms fail decisively. Even when all assumptions intrinsic to the original derivation are incorporated, explanatory power remains negligible. Residual scaling persists only as a weak statistical tendency after normalization by  $Z^2$ , reflecting screening and many-electron effects rather than a predictive law. This failure is therefore not methodological but physical: photon energy alone does not uniquely encode atomic configuration once electron–electron interactions are present.

Orbital angular momentum exhibits intermediate behavior. Exact Coulomb degeneracy suppresses recoverability in hydrogen, while partial recovery emerges in multi-electron systems as angular-momentum-dependent energy shifts break degeneracy. Nevertheless, recovery remains incomplete, and final-state angular momentum collapses into a small number of effective manifolds. Across all models, confusion matrices display pronounced columnar structure, indicating many-to-one spectral mappings rather than model failure. Alternative loss functions and architectural variants modify numerical metrics but do not alter these qualitative outcomes.

Machine-learning models provide an independent empirical probe of this non-invertibility. Without access to atomic identifiers or quantum numbers, neural networks consistently recover initial-state structure while failing to reconstruct final-state quantities. The principal quantum number of the initial state is recoverable at accuracies far exceeding chance, with lower levels showing the strongest fidelity and higher levels collapsing into spectrally compressed manifolds. In contrast, final-state principal quantum numbers are not recoverable under any modeling variant, even when electric dipole selection rules are enforced. This asymmetry reflects the directional loss of information inherent in radiative emission.

A complete element-by-element accounting of recoverability metrics across the periodic table is provided in Appendix 6. Only elements with sufficient post-filter test-set support are included

quantitatively; elements failing this criterion are explicitly marked as insufficient data or collapsed regimes. This ensures that all reported trends reflect genuine spectral information content rather than artifacts of data sparsity or class imbalance.

$Z$	$n_i$	$n_k$	$\ell_i$	$\ell_k$	$\Delta\ell$ mask	$\Delta\ell + n_k < n_i$ mask
1	1.000	0.250	0.550	0.333	0.444	0.000
2	0.941	0.238	0.625	0.357	0.760	0.000
3	0.316	0.282	0.062	0.296	0.629	0.219
4	0.317	0.152	0.338	0.287	0.647	0.000
5	0.242	0.206	0.070	0.275	0.494	0.225
6	0.305	0.294	0.087	0.322	0.606	0.082
10	0.269	0.469	0.217	0.434	0.732	0.005
11	0.372	0.376	0.059	0.294	0.353	0.017
16	0.481	0.294	0.267	0.305	0.653	0.006
18	0.203	0.254	0.337	0.206	0.533	0.000
19	0.560	0.550	0.173	0.516	0.717	0.340
20	0.500	0.791	0.200	0.665	0.882	0.500
22	0.500	0.891	0.333	0.606	0.917	0.000
23	0.750	0.396	0.500	0.349	0.716	0.000
24	0.333	0.692	0.125	0.569	0.753	0.000
25	0.704	0.658	0.327	0.401	0.842	0.000
26	0.322	0.657	0.217	0.548	0.731	0.024
27	0.907	0.594	0.531	0.440	0.939	0.000
29	0.370	0.486	0.242	0.369	0.551	0.189
30	0.960	0.918	0.186	0.476	0.719	0.000
37	0.231	0.025	0.529	0.405	0.692	0.475
38	0.490	0.368	0.070	0.338	0.643	0.036
40	0.250	0.333	0.500	0.500	0.500	0.000
49	0.311	0.027	0.460	0.442	0.706	0.078
50	0.405	0.339	0.132	0.413	0.543	0.286
52	0.500	0.333	0.438	0.464	0.682	0.364
53	0.422	0.584	0.138	0.364	0.590	0.000
55	0.305	0.362	0.118	0.222	0.539	0.310
56	0.304	0.250	0.341	0.150	0.439	0.000
72	0.994	0.333	0.994	0.662	0.988	0.000
78	0.594	0.820	0.343	0.543	0.820	0.451
80	0.191	0.441	0.052	0.393	0.839	0.067
87	0.889	0.833	0.040	0.667	1.000	0.000
88	0.581	0.342	0.222	0.475	0.650	0.000
90	0.639	0.367	0.356	0.487	0.208	0.152

Table 5: Balanced accuracy for recovery of initial and final quantum numbers using the combined spectral model. Mask columns report test-set retention fractions after applying selection rules. Elements with insufficient data in the combined model are excluded from this table.

These results establish a clear hierarchy of information survival under radiative projection. Initial-state structure leaves a detectable imprint on emitted spectra, while final-state structure is largely erased. Recoverability degrades systematically with increasing electronic complexity and spectral degeneracy, reflecting genuine information compression rather than algorithmic limitation.

Neural networks recover precisely this hierarchy without explicit physical guidance, providing empirical confirmation of long-standing principles in atomic physics and clarifying the limits of spectral inversion.

## 10 Conclusion

This study demonstrates that the asymmetry between initial- and final-state predictability reflects a genuine increase in conditional information entropy, in the Shannon sense of reduced mutual information between spectral observables and final-state configuration rather than a deficiency of modeling methodology. When the underlying physical mapping is linear and invertible, as in the Planck–Einstein relation, spectral observables fully encode the relevant structure. In contrast, once atomic configuration enters through conditional, many-body relations, photon energy becomes a many-to-one projection, and inverse reconstruction is fundamentally limited. Machine-learning models recover precisely this hierarchy without explicit physical guidance, empirically delineating the boundary between recoverable and irretrievably lost atomic information. These results clarify the physical limits of spectral inversion and provide a general framework for diagnosing information retention in radiative processes. These findings are fully consistent with established atomic structure theory; their significance lies in the empirical quantification of information loss under controlled observational constraints rather than in contradiction of known physics. The persistence of this hierarchy under architectural variation establishes recoverability itself as a physical observable, not a modeling artifact

## 11 Acknowledgments

The author acknowledges the use of the Pandas [9], TensorFlow [13], Scikit-learn [14], and Matplotlib [15] Python libraries.

The author acknowledges the use of large language model-based assistance (ChatGPT, OpenAI) for code debugging, manuscript structuring, and editorial refinement. All scientific reasoning, analysis, and conclusions are solely those of the author.

The author also acknowledges the support and encouragement of family and close friends during the completion of this work, in particular the author’s spouse, Alexandra.

# References

- [1] Kramida, A. and Ralchenko, Yu. and Reader, J. and NIST ASD Team. Nist atomic spectra database (version 5.11), 2023. National Institute of Standards and Technology.
- [2] Max Planck. On the law of distribution of energy in the normal spectrum. *Annalen der Physik*, 4:553–563, 1901.
- [3] Albert Einstein. On a heuristic point of view concerning the production and transformation of light. *Annalen der Physik*, 17:132–148, 1905.
- [4] Johannes R. Rydberg. On the structure of the line spectra of the chemical elements. *Philosophical Magazine*, 29:331–337, 1890.
- [5] C. J. Foot. *Atomic Physics*. Oxford University Press, 2005.
- [6] Henry G. J. Moseley. The high-frequency spectra of the elements. *Philosophical Magazine*, 26:1024–1034, 1913.
- [7] Robert D. Cowan. *The Theory of Atomic Structure and Spectra*. University of California Press, 1981.
- [8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [9] Wes McKinney et al. pandas: a foundational python library for data analysis and statistics. *Python in Science Conference*, 2010.
- [10] I. I. Sobelman. *Atomic Spectra and Radiative Transitions*. Springer, 1979.
- [11] Eite Tiesinga, Peter J. Mohr, David B. Newell, and Barry N. Taylor. Codata recommended values of the fundamental physical constants: 2018. *Reviews of Modern Physics*, 93:025010, 2021.
- [12] Hans A. Bethe and Edwin E. Salpeter. *Quantum Mechanics of One- and Two-Electron Atoms*. Springer, 1957.
- [13] TensorFlow Developers. Tensorflow, 2024.
- [14] Fabian Pedregosa et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [15] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9:90–95, 2007.



# A Appendix A: Element-wise Recoverability Across the Periodic Table

This appendix summarizes balanced accuracy results for quantum-state recoverability across all elements for which physically consistent preprocessing yielded a non-empty evaluation set. Elements for which filtering collapses the dataset entirely are explicitly marked as non-identifiable, indicating that no inverse problem is defined under spectral-only constraints. Numerical values should be interpreted comparatively rather than absolutely, with emphasis on cross-element trends and asymmetries between initial and final states.

$Z$	$n_i$	$n_k$	$\ell_i$	$\ell_k$	$\Delta\ell$ mask	$\Delta\ell + n_k < n_i$ mask	Comment
1	1.000	0.250	0.550	0.333	0.444	0.000	OK
2	0.620	0.454	0.244	0.326	0.619	0.000	OK
3	0.316	0.282	0.062	0.296	0.629	0.219	OK
4	0.317	0.152	0.338	0.287	0.647	0.000	OK
5	0.242	0.206	0.070	0.275	0.494	0.225	OK
6	0.305	0.294	0.087	0.322	0.606	0.082	OK
7	—	—	—	—	—	—	Insufficient data
8	—	—	—	—	—	—	Insufficient data
9	—	—	—	—	—	—	Insufficient data
10	0.269	0.469	0.217	0.434	0.732	0.005	OK
11	0.372	0.376	0.059	0.294	0.353	0.017	OK
12	—	—	—	—	—	—	Insufficient data
13	—	—	—	—	—	—	Insufficient data
14	—	—	—	—	—	—	Insufficient data
15	—	—	—	—	—	—	Insufficient data
16	0.481	0.294	0.267	0.305	0.653	0.006	OK
17	—	—	—	—	—	—	Insufficient data
18	0.203	0.254	0.337	0.206	0.533	0.000	OK
19	0.560	0.550	0.173	0.516	0.717	0.340	OK
20	0.500	0.791	0.200	0.665	0.882	0.500	OK
21	0.500	0.122	0.500	0.406	0.550	0.000	Low support
22	0.500	0.891	0.333	0.606	0.917	0.000	OK
23	0.750	0.396	0.500	0.349	0.716	0.000	OK
24	0.333	0.692	0.125	0.569	0.753	0.000	OK
25	0.704	0.658	0.327	0.401	0.842	0.000	OK
26	0.322	0.657	0.217	0.548	0.731	0.024	OK
27	0.907	0.594	0.531	0.440	0.939	0.000	OK
28	—	—	—	—	—	—	Insufficient data
29	0.370	0.486	0.242	0.369	0.551	0.189	OK
30	0.960	0.918	0.186	0.476	0.719	0.000	OK

$Z$	$n_i$	$n_k$	$\ell_i$	$\ell_k$	$\Delta\ell$ mask	$\Delta\ell + n_k < n_i$ mask	Comment
31	—	—	—	—	—	—	Insufficient data
32	—	—	—	—	—	—	Insufficient data
33	—	—	—	—	—	—	Insufficient data
34	—	—	—	—	—	—	Insufficient data
35	—	—	—	—	—	—	Insufficient data
36	—	—	—	—	—	—	Insufficient data
37	0.231	0.025	0.529	0.405	0.692	0.475	OK
38	0.490	0.368	0.070	0.338	0.643	0.036	OK
39	—	—	—	—	—	—	Collapsed (low support)
40	0.250	0.333	0.500	0.500	0.500	0.000	Collapsed (low support)
41	—	—	—	—	—	—	Insufficient data
42	—	—	—	—	—	—	Insufficient data
43	—	—	—	—	—	—	Insufficient data
44	—	—	—	—	—	—	Insufficient data
45	—	—	—	—	—	—	Insufficient data
46	—	—	—	—	—	—	Insufficient data
47	—	—	—	—	—	—	Low support (n/l only reliable)
48	—	—	—	—	—	—	Insufficient data
49	0.311	0.027	0.460	0.442	0.706	0.078	OK
50	0.405	0.339	0.132	0.413	0.543	0.286	OK
51	—	—	—	—	—	—	Insufficient data
52	0.500	0.333	0.438	0.464	0.682	0.364	OK
53	0.422	0.584	0.138	0.364	0.590	0.000	OK
54	—	—	—	—	—	—	Insufficient data
55	0.305	0.362	0.118	0.222	0.539	0.310	OK
56	0.304	0.250	0.341	0.150	0.439	0.000	OK
57	—	—	—	—	—	—	Insufficient data
58	—	—	—	—	—	—	Insufficient data
59	—	—	—	—	—	—	Insufficient data
60	—	—	—	—	—	—	Insufficient data
61	—	—	—	—	—	—	Insufficient data
62	—	—	—	—	—	—	Insufficient data
63	—	—	—	—	—	—	Insufficient data
64	—	—	—	—	—	—	Insufficient data
65	—	—	—	—	—	—	Insufficient data
66	—	—	—	—	—	—	Insufficient data
67	—	—	—	—	—	—	Insufficient data
68	—	—	—	—	—	—	Insufficient data
69	—	—	—	—	—	—	Insufficient data
70	—	—	—	—	—	—	Insufficient data
71	—	—	—	—	—	—	Insufficient data
72	0.994	0.333	0.994	0.662	0.988	0.000	Collapsed (near-deterministic)
73	—	—	—	—	—	—	Insufficient data
74	—	—	—	—	—	—	Collapsed (all-ones artifact)
75	—	—	—	—	—	—	Insufficient data

$Z$	$n_i$	$n_k$	$\ell_i$	$\ell_k$	$\Delta\ell$ mask	$\Delta\ell + n_k < n_i$ mask	Comment
76	—	—	—	—	—	—	Insufficient data
77	—	—	—	—	—	—	Collapsed (canon-only)
78	0.594	0.820	0.343	0.543	0.820	0.451	OK
79	—	—	—	—	—	—	Insufficient data
80	0.191	0.441	0.052	0.393	0.839	0.067	OK
81	—	—	—	—	—	—	Insufficient data
82	—	—	—	—	—	—	Insufficient data
83	—	—	—	—	—	—	Insufficient data
84	—	—	—	—	—	—	Insufficient data
85	—	—	—	—	—	—	Insufficient data
86	—	—	—	—	—	—	Insufficient data
87	0.889	0.833	0.040	0.667	1.000	0.000	Low support
88	0.581	0.342	0.222	0.475	0.650	0.000	OK
89	—	—	—	—	—	—	Insufficient data
90	0.639	0.367	0.356	0.487	0.208	0.152	OK
91	—	—	—	—	—	—	Insufficient data
92	—	—	—	—	—	—	Insufficient data

Table 6: Full elemental sweep of balanced accuracies for quantum-state recovery using the combined spectral model. Elements marked as *Insufficient data* lacked adequate test-set support after filtering. Rows marked as *Collapsed* exhibit deterministic or near-deterministic metrics indicative of trivial or underconstrained classification regimes and are excluded from quantitative interpretation.