

Redshift estimation on astronomy spectral data using CNN-based architecture with correction*

Iryna Yurchuk¹ and Vladyslav Skyruta^{1,*}

¹ Taras Shevchenko National University of Kyiv, Volodymyrska str. 64/13, Kyiv, 01601, Ukraine

Abstract

Redshift is a fundamental characteristic of cosmic objects, particularly galaxies, stars, and quasars. By measuring the redshift of a galaxy, it is possible to determine its distance, velocity, and the expansion of the universe, as well as study the relative motion of cosmic structures. For investigating objects with high redshifts – where visible light is shifted into the infrared spectrum – scientists rely on advanced orbital telescopes, such as the James Webb Space Telescope.

In this study, an efficient algorithm for estimating the redshift of galaxies based on their spectra has been developed. The algorithm has an architecture built upon CNNs, and incorporates a correction mechanism for the final results using laboratory-calibrated positions of key emission lines in galaxy spectra.

Trained on a dataset of 10,000 samples consisting of starburst galaxies with redshifts ranging from 0 to 0.6, the model achieves a mean absolute error of 0.0169 without and 0.0139 with correction. This corresponds to an average relative accuracy of approximately 97.1% without and 97.7% with correction. The proposed method demonstrates significant potential for improving the accuracy and efficiency of redshift estimation in large-scale astronomical surveys.

Keywords

redshift estimation, spectral analysis, emission lines, machine learning

1. Introduction

The study of redshift is a cornerstone of modern astrophysics and cosmology. It provides critical insights into the structure, dynamics, and evolution of the universe. High-redshift objects, such as distant galaxies and quasars, are particularly valuable for understanding the early stages of cosmic evolution [1]. However, accurately estimating redshift from galaxy spectra remains a challenging task, especially for large datasets generated by modern astronomical surveys, such as SDSS, Euclid, NASA, etc. In this context, redshift estimation stays as the central research object of the study.

And this is where machine learning algorithms are jumping in. Machine learning is good at handling large and complex datasets. It is an ideal tool for modern astronomical research, where huge amounts of spectral data are generated by large-scale surveys and advanced telescopes ([2] and [3]). These algorithms are particularly good at identifying complicated patterns and correlations within data that may be difficult or impossible to distinguish using traditional analytical methods. As such, the application of machine learning techniques to galaxy spectra serves as the research subject explored in this work.

There have already been numerous studies that apply machine learning techniques to the problem of redshift estimation. Many of these approaches utilize traditional machine learning models, such as Random Forests or Support Vector Machines, as well as more recent deep learning architectures like standard CNNs ([4] and [5]). These methods have demonstrated significant success in automating redshift estimation and improving its accuracy compared to classical techniques. However, they

*International Workshop on Computational Intelligence, co-located with the IV International Scientific Symposium "Intelligent Solutions" (IntSol-2025), May 01-05, 2025, Kyiv-Uzhhorod, Ukraine

^{1*} Corresponding author.

✉ i.a.yurchuk@gmail.com (I. Yurchuk); vlad.skyruta@gmail.com (V. Skyruta)

ORCID 0000-0001-8206-3395 (I. Yurchuk); 0009-0000-2599-3924 (V. Skyruta)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

often face limitations in terms of scalability, computational efficiency and their ability to generalize features across diverse datasets.

In contrast, the approach presented in this work make use of advanced Inception network architecture, which is specifically designed to capture multi-scale features within data. By incorporating a correction mechanism based on laboratory-calibrated positions of key emission lines, our method achieves higher precision in redshift estimation. This combination of advanced architecture and post-processing refinement allows the algorithm to outperform existing methods, particularly when applied to large datasets of galaxies with varying spectral characteristics. The purpose of this study is to refine redshift estimation techniques by enhancing accuracy and scalability for large astronomy datasets.

2. Theoretical Background and Related Work

2.1. Theory and concepts

When a beam of white light passes through a triangular prism, it disperses into its fundamental components, known as the spectrum. Spectra play crucial role in modern astronomical research, since they allow scientists to determine the qualitative and quantitative chemical composition of celestial objects, their temperature, the presence of magnetic fields, their velocity and much more.

As light leaves the surface of a cosmic object, it passes through its atmosphere. During this process, certain wavelengths are absorbed by the chemical elements present in the atmosphere. This absorption results in gaps or dark lines in the spectrum, which are known as Fraunhofer absorption lines.

In contrast to absorption lines, emission lines are bright lines that appear in the spectrum when atoms or molecules emit light at specific wavelengths. These lines are produced when electrons in atom transition from a higher energy level to a lower one, releasing energy in the form of photons. Emission lines are typically observed in regions of hot, ionized gas, such as nebulae or the outer layers of certain stars.

Each chemical element has a unique spectral signature, consisting of specific absorption or emission lines. In contrast, astronomical spectra, for example of galaxies, are complex combinations of the spectra from numerous sources, including stars, gas clouds, and dust. These spectra include a continuous background, absorption and emission lines.

By analyzing both absorption and emission lines of astronomical spectra, astronomers can gain a comprehensive understanding of the physical and chemical properties of celestial objects. For instance, they can easily identify the chemical elements that absorbed or emitted the light, thereby determining the chemical composition of the atmosphere or surrounding gas. Since the laws of physics are assumed to operate similarly throughout the Universe, it is expected that the absorption and emission lines of common chemical elements should appear the same everywhere.

However, when astronomers observe the spectra of celestial objects, they almost always find that their spectral lines are “shifted” toward the red (longer wavelength) end of the spectrum. This phenomenon is known as the redshift.

The redshift of light can occur due to several well-known phenomena. The most common cause is the Doppler effect, where light from an object moving away from the observer is stretched to longer wavelengths. Another significant cause is the cosmological redshift, which arises from the expansion of the Universe, stretching the wavelengths of light as it travels through space. Additionally, gravitational redshift occurs when light escapes a strong gravitational field, losing energy and shifting to longer wavelengths.

Redshift is a fundamental concept in astronomy. By observing how the color of a celestial object changes over time, or how it differs from what was expected, researchers can discover a lot of

interesting facts about the object. Redshift quantifies how much the wavelength of electromagnetic radiation detected on Earth has changed compared to the wavelength emitted by the source.

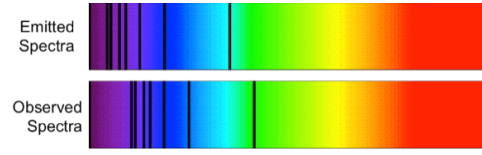


Figure 1: Demonstration of shifted absorption lines (<https://voyages.sdss.org>)

It is, of course, impossible to physically capture the light beam of a distant star before it has been redshifted during its journey to Earth. Therefore, astronomers use alternative methods to measure redshift, the primary ones being the spectrographic and photometric methods.

Spectrographs are highly powerful tools for studying light from cosmic objects. They split light into its constituent colors, allowing scientists to analyze each color range in detail. By comparing the absorption lines of the most common chemical elements in the Universe with their standard values, astronomers can determine how much the entire spectrum has been “shifted” in comparison to its original state when the light left the celestial body. This approach to measure redshift is called spectrographic method.

Photometers, on the other hand, measure the total intensity of light within a specific range of frequencies. While this method is less detailed and affects the accuracy of redshift calculations – since it considers only averaged values of the spectrum rather than its full details – it is faster and simpler.

The photometric method of estimating redshift is particularly useful when analyzing large clusters of cosmic objects, where precise measurements are not mandatory. However, when astronomers conduct detailed studies of individual objects, they rely on spectral data and the spectrographic method for determining redshift.

2.2. Review of Existing Approaches

First of all, there is a significant number of studies that have been conducted to estimate photometric redshift. Usually they take advantage of using traditional machine learning methods, as in [6] or [7]. Data Release 17 of SDSS also use kd-tree nearest neighbor fit, described in [8], to estimate photometric redshifts efficiently. However, there are also plenty of examples of applying deep learning techniques [9].

Despite these features, photometric redshift estimation remains challenging due to the limitations of photometric data. The lack of detailed spectral information often introduces uncertainties and biases, which can impact the accuracy of the calculated redshifts.

On the other hand, some studies, focus on calculating redshift using spectral data. This approach is often considered more accurate than photometric methods, because spectral data provides detailed information about the light emitted by celestial objects across a wide range of wavelengths, unlike photometric methods, which rely on broad-band magnitudes.

Since CNNs are good at extracting features from 1D data, there have been already conducted some researches using them. For instance, in [10] besides the fact that authors transform the redshift estimation problem from regression to classification by dividing the redshift range into discrete intervals based on Euclid spectroscopic instrument, they use CNNs to classify galaxies into these intervals as well.

In addition to that, there is a study that utilize Bayesian CNNs inspired by VGG architectures to estimate spectroscopic redshift [11]. It provides both predictions and uncertainty estimates, enabling the identification of problematic spectra and balancing prediction accuracy with coverage. The method outperforms traditional template-fitting techniques and helps discover misclassified or unrecognized quasars.

3. Proposed Method

The proposed approach aims to extract as much useful information as possible from spectral data to achieve high-precision redshift estimation. Since spectra are essentially 1D data, CNNs are well-suited for analyzing them. In this work, we propose not only using a standard one- or multi-layer CNN for feature extraction but also incorporating the Inception network [12].

The Inception network is designed to capture features at multiple scales by applying convolutional filters of different sizes in parallel. This architecture potentially is advantageous for spectral data, as it allows the model to simultaneously detect both narrow features, such as emission or absorption lines, and broader patterns, such as continuum variations.

In addition to the primary model, an additional correction method is introduced to further refine the redshift estimation. After the initial redshift is predicted by the main model, the spectrum is “shifted” back to its born state based on the calculated redshift. In this initial state, emission lines are basically aligned with their expected laboratory-calibrated positions. Using these corrected spectra, the model adjusts its predictions by recalculating the redshift based on the ideal positions of the emission lines and considering spectra characteristics, such as peaks and flat regions. This additional correction step ensures more precise redshift value.

3.1. Data preparation

3.1.1. Key notes about the dataset

The model will be evaluated on Data Release 18 of SDSS [13], the latest and most comprehensive dataset from the Sloan Digital Sky Survey, offering high-quality spectral data across a wide range of celestial objects.

Since stars, QSOs, and galaxies exhibit fundamentally different features in their spectra, this study focuses only on galaxy spectra with redshifts up to 1. It is worth mentioning that the proposed approach can be adapted for stars and quasars as well, making it flexible for broader applications.

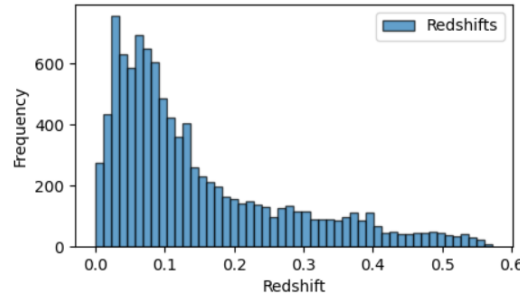


Figure 2: Redshift distribution over the dataset

To ensure a homogeneous dataset, only starburst galaxies are to be considered in this work. Starburst galaxies are characterized by a recent and transient increase in their SFR, often by a factor of up to 50 compared to regular galaxies. The strong radio emissions of starburst galaxies make it easier for machine learning models to extract meaningful features from their spectra, thus enhancing the reliability and precision of redshift estimation. It is expected that the proposed approach will also work for regular galaxies, though this would require reconsidering the data preprocessing methods described in the next subsection.

3.1.2. Models inputs

First of all, let's define the structure of the input data and highlight its key features relevant to this work. Since there are two models involved – the main model and the correction model – each

requires a distinct set of inputs. However, both models share a common input – the spectrum, which represents the flux values (in 10^{-17} erg/s/cm²/Å) corresponding to specific wavelengths (in angstroms Å).

Typically, galaxy spectra provide flux information within the range of approximately 4000 to 9000 angstroms, as this range covers key spectral features such as the Balmer series, prominent emission lines and other important indicators of galaxy properties. This range is also well-suited for optical observations, as it falls within the sensitivity of most ground-based telescopes and spectrographs.

In addition to the spectrum, the main model requires the redshift value as an input. Redshift, denoted as z , is defined as the fractional change in the wavelength of light due to various physical phenomena. It can be calculated using the formula

$$z = \frac{\lambda_{obs} - \lambda_{rest}}{\lambda_{rest}}, \quad (1)$$

where λ_{obs} is a wavelength measured by the observer, and λ_{rest} is the wavelength at the source.

The correction model, on the other hand, takes as input the spectra shifted back to their emitted states, along with the template positions of several well-known and widely distributed emission lines (e.g. H α , H β , [OIII], [NII]). After shifting the spectra to their source states, any empty values that arise due to the shift are filled with zeros, while values that fall outside the defined above wavelength range are removed.

3.1.3. Data preprocessing

Since spectra are 1D signals often containing noise from various sources, wavelet filtering is an effective solution for denoising while preserving important features. Wavelet filters are a well-established tool for noise reduction and are widely used in various applications. For instance, in [15], wavelet filtering is applied to detect anomalies and emergency states in signals by isolating key characteristics after removing noise.

For noise reduction, the Daubechies wavelet is commonly used due to its ability to capture smooth and localized features in 1D signals. The filtering process involves decomposing the signal into wavelet coefficients across multiple scales, where noise is typically concentrated in the high-frequency components. A soft thresholding technique is applied to suppress these noisy coefficients, ensuring that the essential features of the signal are preserved. After thresholding, the signal is reconstructed by combining the modified coefficients, resulting in a cleaner version of the original data.

It's worth mentioning that there is no need to completely remove all features from the spectra, as some degree of noise is acceptable if it preserves the essential characteristics of the signal. The chosen threshold, set to 10% of the maximum coefficient at the finest scale, keeps a balance between effective noise reduction and retaining meaningful spectral features, as shown on Figure 3. This value ensures that weaker but important components, such as emission lines, are not completely suppressed, while still keeping the majority of high-frequency noise.

The next step in the data preprocessing pipeline involves applying spline approximation to the spectrum. This technique is very useful for starburst galaxies, whose spectra are characterized by a relatively flat continuum with distinct emission lines standing out prominently.

Then, by calculating the difference between the wavelet-filtered spectrum and the spline approximation line, we align the spectrum with the zero axis, removing base flux variations and focusing on relative flux changes (Figure 4). While this process sacrifices some physical context, it is highly beneficial for further training the convolution-based networks, as they are better suited to analyze relative variations rather than absolute flux values. Finally, the resulting difference is normalized by dividing it by its maximum value, ensuring consistent scaling across all spectra and preparing the data for efficient feature extraction and model training.

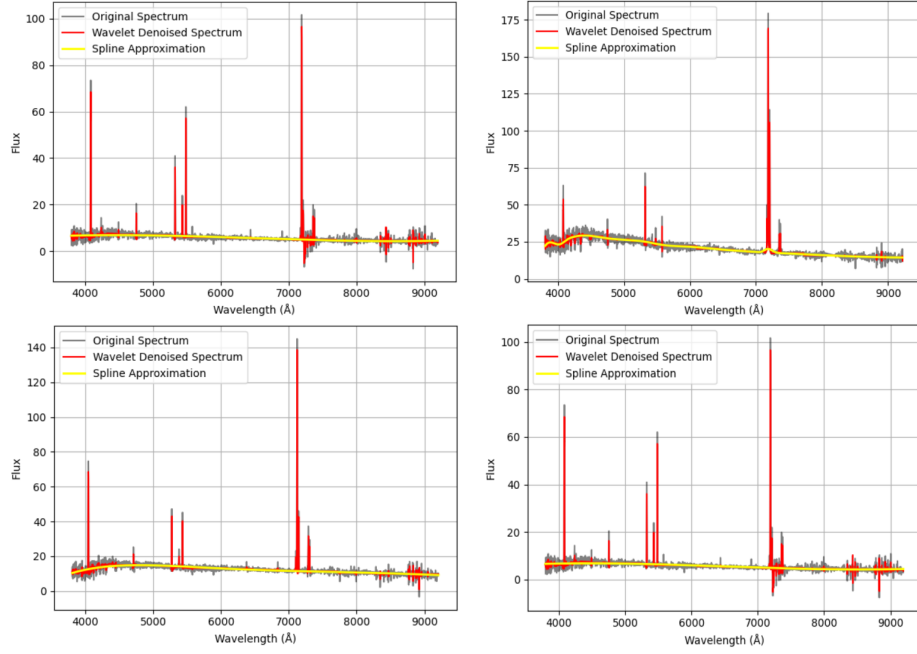


Figure 3: Examples of applying Daubechies wavelet to the spectra and estimating approximation spline

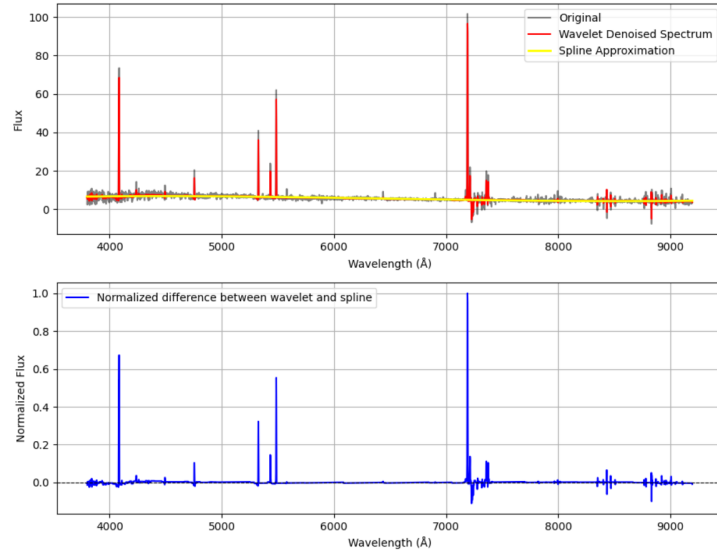


Figure 4: Illustration of spectrum filtering. The final result, along with other spectrum characteristics, is passed to the machine learning model

3.2. Building the machine learning model

As mentioned earlier, 1D CNNs are widely used for processing sequential data, such as time series or audio signals, to detect important patterns or identify unwanted anomalies. Given the sequential nature of astronomical spectra, they are excellent candidates for extending the scope of 1D CNN applications to the field of astronomy.

In general, CNNs (Convolutional Neural Networks) are a class of deep learning models designed to automatically and adaptively learn spatial hierarchies of features from data. The core operation in CNNs is the convolution, which involves applying a filter to the input data to extract features. Mathematically, a convolution operation for a 1D input can be expressed as

$$y(i) = \sum_j x(j) * w(i-j), \quad (2)$$

where $x(j)$ is the input signal, $w(i-j)$ is a filter, and $y(i)$ is the resulting feature map.

A 1D CNN is a specialized type of CNN where the convolution operation is applied along a single dimension.

One of the enhancements of CNN demonstrated in this work is the Inception network architecture [12]. The main idea behind it is to process input data at multiple scales simultaneously by using parallel convolutional layers with different kernel sizes.

The Inception architecture was chosen for redshift prediction due to its ability to detect in multiple scales both narrow spectral features, such as emission lines, and broader patterns, such as continuum variations. Additionally, its modular design allows for efficient computation.

3.3. Correction method

The correction method is designed to further enhance the accuracy of the overall algorithm. In regression tasks, convolutional networks typically return a single redshift value, which is not directly tied to specific spectral features, such as emission or absorption lines. To address this limitation, we propose a correction mechanism that aligns the predicted redshift value with the spectral characteristics.

But, in the first place it is necessary to calculate the positions of emission lines based on the estimated redshift value using Formula 1. These calculated positions allow us to adjust the emission lines, aligning them more closely with the observed spectrum peaks, thereby refining the redshift estimation.

To begin with, the most naïve method to adjust emission line positions is a greedy algorithm. The idea behind this approach is to iteratively refine the predicted positions of emission lines by searching for the best match within a local region of the spectrum.

For each predicted position, the algorithm evaluates nearby points within a defined search radius, considering three main factors:

- proximity to the initial prediction;
- intensity of the spectrum at the point;
- relative distance between neighboring emission lines.

The algorithm assigns a cost to each candidate position based on these criteria and selects the one with the lowest cost. This process ensures that the corrected positions align more closely with the actual peaks in the spectrum while maintaining consistent spacing between emission lines.

Another approach to refine emission line positions involves training a machine learning model using local regions of the spectrum. For that, gradient boosting method is used. It is an ensemble learning algorithm that builds a series of decision trees. They can be trained to solve either classification or regression tasks. In our case, the regressor is built to predict positions of emissions lines.

Precise estimation of all mentioned algorithms will be presented in the next section.

4. Experimental Evaluation

Important, that the input spectra do not share a common wavelength grid, which complicates direct comparisons and processing. Each spectrum is sampled at different wavelengths, making it necessary to align all spectra to a unified grid before further analysis. To fix this issue, a common wavelength grid was defined in the range from 4000 Å to 9000 Å with step equal to 1 Å. The flux values are interpolated onto the common grid using a linear interpolation method. Chosen grid has direct impact on the model performance and accuracy. A finer grid with smaller step sizes keeps more

precise representation of spectral features, improving the model performance. At the same time, increased accuracy comes at the cost of higher computational complexity and memory usage, as the model must process a larger number of data points.

4.1. CNN model evaluation

The evaluation of the algorithm begins with testing the main model, which is based on a CNN. For this purpose, a single-layer CNN was implemented. The performance of the model varies depending on the number of the convolutional filters and their sizes, as shown in Table 1.

Table 1
Single layer CNN model accuracy

Number of filters	Filter size	Redshift MAE	Relative accuracy, %
64	3	0.0322	94.35
64	150	0.0169	97.03
64	350	0.0173	96.96
128	3	0.0286	94.98
128	150	0.0182	96.80
128	350	0.0204	96.42

Trainings were conducted on the CNN using the ReLU activation function and the mean absolute error loss function over 20 epochs. In addition to convolution filter, average pooling is applied with filter size of 2.

As it is shown in the Table 1, smaller convolution filters result in higher MAE values since they are less effective at capturing broader spectral features. On the other hand, larger filters demonstrate improved accuracy, with the lowest MAE achieved using 64 filters and a filter size of 150. Increasing the number of filters to 128 does not consistently improve performance. While increasing the number of filters can enhance feature extraction, it may also introduce redundancy or overfitting, depending on the filter size.

To achieve a more precise estimation of redshift, the Inception network was used. Since the model's complexity and computational efficiency are critical for handling large datasets, the architecture with dimensionality reduction was used (Figure 5). It consists of three to four parallel towers: two or three convolutional filters of different sizes applied after a 1x1 convolution, and a max pooling operation followed by another 1x1 convolution.

Trainings was conducted on the Inception network consisting of three towers using the ReLU activation function again as well as the mean absolute error loss function over 20 epochs. Results can be found in Table 2.

Considering the results, the Inception network demonstrates significantly higher MAE values compared to the single-layer CNN model. The lowest MAE achieved by the Inception network is 0.0504, which is worse than the best result of 0.0169 obtained by the CNN with 64 filters and a filter size of 150. This imbalance can be explained by the simplicity and efficiency of the CNN architecture, which is better suited for the straightforward task of redshift estimation. In contrast, the Inception network's complex architecture may introduce unnecessary computational overhead and fail to capture the specific features required for accurate redshift prediction.

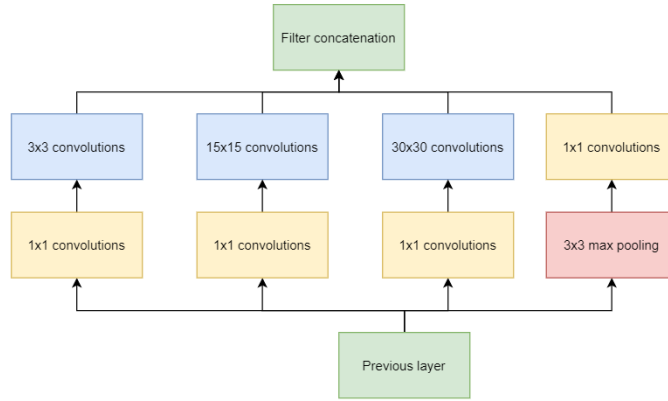


Figure 5: Architecture of used Inception network

Table 2

Inception network model accuracy

Number of filters	Convolutions	Pooling	Redshift MAE	Relative accuracy, %
64	3x3 15x15	2x2	0.0504	91.00
64	20x20 50x50	2x2	0.0861	84.62
128	3x3 15x15	2x2	0.0953	82.98

4.2. Correction method evaluation

Since there is no difference which CNN architecture choose to test correction approach, single layer CNN with 64 filters and a filter size of 150 will be used further.

The CNN model outputs the estimated redshift value for a given spectrum. It is then used to calculate the positions of emission and absorption lines. These calculated positions are not aligned with specific spectral characteristics, such as peaks, as shown on Figure 6.

As described earlier, two approaches are going to be compared – the gradient boosting model and the greedy algorithm.

The gradient boosting is trained using 100 estimator trees and a 0.1 shrinkage parameter. Mean squared error is a loss function. The greedy algorithm assigns a cost to each candidate position based on its proximity to the initial prediction, the intensity of the wavelength at the point, and how well the new position maintains the spacing between neighboring emission lines. The positions with the lowest cost are selected as the new candidates for the emission lines.

Eight emission lines are chosen for performing correction: H β , [O III] 4959, [O III] 5007, [N II] 6548, H α , [N II] 6583, [S II] 6716, [S II] 6730.

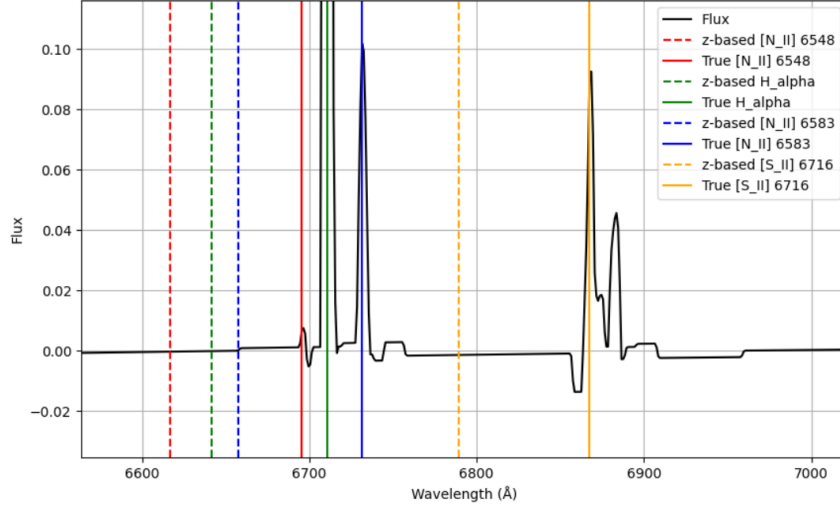


Figure 6: Illustration of how emissions lines calculated based on the estimated redshift are located in comparison to their true locations. It can be noticed that the true positions align with the most visible peaks in the spectrum, since they correspond to light emission at specific wavelength. On the other hand, the calculated positions are often located in a flat regions of the spectrum. The figure includes a zoomed-in section of the spectrum for better visualization

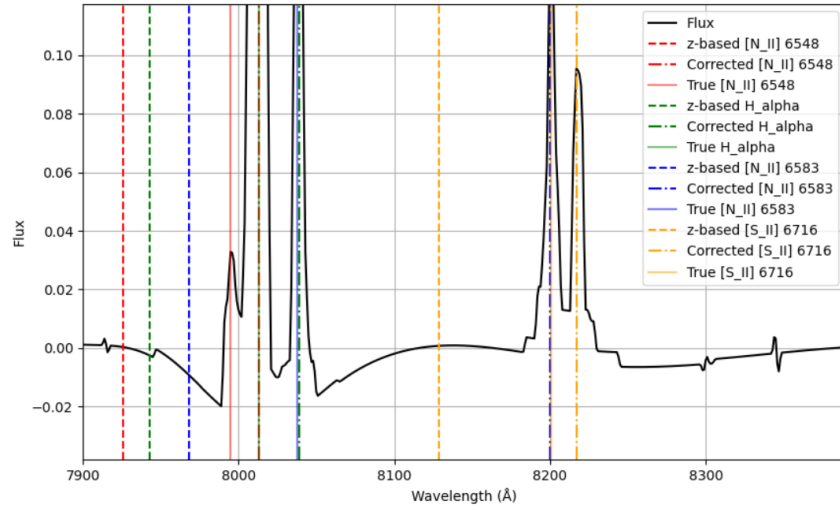


Figure 7: Example of emission line positions correction using greedy algorithm. For [N II] 6548, estimated position is at 7925 Å, error – 69 Å, corrected – at 8012 Å, true – at 7994 Å, error after correction – 18 Å. For H α , estimated position is at 7942 Å, error – 70 Å, corrected – at 8038 Å, true – at 8012 Å, error after correction – 26 Å. For [N II] 6583, estimated position is at 7967 Å, error – 70 Å, corrected – at 8199 Å, true – at 8037 Å, error after correction – 162 Å. For [S II] 6716, estimated position is at 8128 Å, error – 72 Å, corrected – at 8216 Å, true – at 8200 Å, error after correction – 16 Å

Table 3 contains the results of the correction model using different approaches. The region size indicates the size of the local window around each emission line where adjustments can be made. It can be observed from the table that both the greedy method and gradient boosting show improvements in redshift prediction accuracy as the region size increases. These results show that both methods are effective, however still gradient boosting provides more consistent corrections, especially for larger region sizes.

Table 3
Results of correction model

Region size	Greedy method MAE	Gradient boosting MAE	Initial accuracy improvement (greedy), %	Initial accuracy improvement (GB), %
150	0.016603	0.014634	1.76	13.41
300	0.015094	0.014439	10.69	14.56
450	0.014333	0.014330	15.19	15.21
600	0.013843	0.014142	18.09	16.32
750	0.013898	0.013963	17.76	17.38

5. Conclusions

This paper describes the effective approach of redshift estimation using spectral data and machine learning techniques. The proposed algorithm combines convolutional neural networks (CNNs) for initial redshift prediction with correction methods designed to refine the estimated values based on spectral characteristics.

The final results of the model achieve up to 97.7% accuracy in redshift estimation when utilizing gradient boosting correction.

Future work could explore hybrid approaches or further optimization of the Inception network to better align its capabilities with the requirements of redshift estimation. Additionally, the correction methods could be enhanced by incorporating more sophisticated spectral features or involving advanced machine learning techniques, such as attention mechanisms, to better capture the relationships between emission lines and spectral peaks.

Overall, the combination of CNN-based redshift estimation and correction methods provides a promising framework for accurate and efficient analysis of spectral data.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

- [1] Haojing Yan et al, First Batch of $z \approx 11-20$ Candidate Objects Revealed by the James Webb Space Telescope Early Release Observations on SMACS 0723-73, *ApJL* 942 L9 (2023). doi:10.3847/2041-8213/aca80c.
- [2] M. Jiménez, M. Torres Torres, R. John and I. Triguero, Galaxy Image Classification Based on Citizen Science Data: A Comparative Study, *IEEE Access* 8 (2020), 47232-47246. doi:10.1109/ACCESS.2020.2978804.
- [3] Abel Dana Ruiz: Detection of Gravitational Wave signals using Machine Learning methods and Generative Pre-trained Transformers, Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.
- [4] Y. Zhang, X. Jin, J. Zhang and Y. Zhao, Machine Learning for Photometric Redshift Estimation of Quasars with Different Samples, in: 2020 IEEE International Conference on Visual

- Communications and Image Processing, Macau, China, 2020, pp. 294-297, doi:10.1109/VICIP49819.2020.9301849.
- [5] A. D’Isanto and K. L. Polsterer, Photometric redshift estimation via deep learning, *A&A* 609, A111 (2018). doi:10.1051/0004-6361/201731326.
 - [6] Hu Zou et al, Photometric Redshifts and Galaxy Clusters for DES DR2, DESI DR9, and HSC-SSP PDR3 Data, *Res. Astron. Astrophys.* 22 065001 (2022). doi:10.1088/1674-4527/ac6416.
 - [7] Vikhyat Agarwal et al, SpyderZ: An Efficient Support Vector Machine Library for Photometric Redshift Estimation and Redshift Probability Information, in *Res. Notes AAS* 8 126 (2024). doi:10.3847/2515-5172/ad47fb.
 - [8] Csabai, I., Multidimensional indexing tools for the virtual observatory, *Astronomische Nachrichten*, vol. 328, no. 8, p. 852, 2007. doi:10.1002/asna.200710817.
 - [9] Q. Lin, D. Fouchez, J. Pasquet, M. Treyer, R. Ait Ouahmed, S. Arnouts, O. Ilbert, Photometric Redshift Estimation with Convolutional Neural Networks and Galaxy Images: A Case Study of Resolving Biases in Data-Driven Methods. *A&A* 662, A36 (2022). doi:10.1051/0004-6361/202142751.
 - [10] R. Stivaktakis, G. Tsagkatakis, B. Moraes, F. Abdalla, J. -L. Starck and P. Tsakalides, Convolutional Neural Networks for Spectroscopic Redshift Estimation on Euclid Data, in *IEEE Transactions on Big Data*, vol. 6, no. 3, pp. 460-476, 2020, doi:10.1109/TBDDATA.2019.2934475.
 - [11] O. Podsztaev, P. Škoda, P. Tvrdík, Spectroscopic redshift determination with Bayesian convolutional networks, *Astronomy and Computing* 40 100615 (2022). doi:10.1016/j.ascom.2022.100615.
 - [12] Szegedy, Christian: Going deeper with convolutions, in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1-9, 2015. doi:10.1109/CVPR.2015.7298594.
 - [13] Andrés Almeida et al, The Eighteenth Data Release of the Sloan Digital Sky Surveys: Targeting and First Spectra from SDSS-V, *ApJS* 267 44 (2023). doi:10.3847/1538-4365/acda98.
 - [14] Ivana Orlitova, Starburst galaxies, in *Reviews in Frontiers of Modern Astrophysics: From Space Debris to Cosmology*; Springer 2020. doi:10.1007/978-3-030-38509-5.
 - [15] Olha Oliynyk, Yuri Taranenko, Valerii Lopatin, Analysis of Discrete Wavelet Spectra of Broad Signals, *Computer Modeling and Intelligent Systems*, 3392, 188-198. doi:10.32782/cm/s/3392-16.
 - [16] Carter L Rhea et al, A machine learning approach to galactic emission-line region classification, *RAS Techniques and Instruments*, vol. 2, issue 1, 345–359. doi.org:10.1093/rasti/rzad023.