ESTIMATING THE DISTANCE OF GALAXIES FROM IMAGES AND MAGNITUDES

Anichebe K Osita under the supervision of Dr Daniel Farrow

Department of Artificial intelligence and Data science, The University of Hull, Cottingham Road, Hull City, United Kingdom.

Abstract

Determining accurate distances of galaxies is important for unraveling the cosmological structure, evolution, and dynamics of the universe on large scales. Galaxies span an immense range of distances from our closest, the Andromeda Galaxy which is 2.537 million light years from earth to the most distant galaxy yet discovered named HD1 of some 13.5 billion light years away (Buongiorno, 2023). Precisely measuring their distances enables mapping the 3D distribution of matter over cosmic time to understand their structure formation and expansion history.

The present standard for galaxy distance measurement is by determining spectroscopic redshifts (z) from the shift in spectral lines due to cosmic expansion. As cosmology has entered an era of massive data and information (Ishida 2019), obtaining spectra for the billions of galaxies visible to modern surveys is infeasible. Photometric redshifts present a convenient and efficient alternative approach to estimating galaxy distances using their multi-band photometric measurements. By sampling a galaxy's spectral energy distribution (SED) in different optical and near-infrared filters, the redshift and hence, distance can be deduced.

This project undertakes a comprehensive investigation into improving the accuracy and reliability of photometric redshift estimation using a multi-modal machine learning approach. The study utilizes both image data from galaxy surveys as well as catalogs of photometric band magnitudes (flux information) to develop a redshift prediction model. Deep learning image analysis can capture detailed morphological indicators related to redshift, complementing the color and magnitude inputs from photometry which leads to better performance compared to using photometry alone. This is evidenced by the lower RMSE (root mean squared error) of 0.019484 and higher R2 of 0.968 for the multi-modal model, compared to 0.0543 and 0.761 for photometry obtained in this study.

By leveraging the strengths of both image analysis and photometry data, this project aims to push the boundaries of photometric redshift precision and robustness by utilizing a multimodal model structure at its core.

Introduction and Background

Redshift is a phenomenon detailing the state of electromagnetic radiation (such as light) from an object that has undergone an increase in wavelength (Jarvis, 2020). There are three main causes of redshift which is attributed to the following:

- Doppler redshift - Light from moving objects will appear to have different wavelengths depending on the relative motion of the source and the observer. When their positions is receding, the wavelength increases and the color shifts towards the red-light spectrum (British Astronomical Association, 2022).

- The cosmic redshift – This result from the expanding universe, known as the metric expansion of space (Whiting, 2004), causing objects to become more distant without changing their more local positions in space. The expansion of space causes light from distant galaxies to stretch to longer wavelengths as the galaxies move away from us (Hubble, 1929). This is the dominant cause of redshift we observe from other galaxies.

- Gravitational redshift - Light escaping a gravitational field loses energy and is shifted to longer wavelengths (Rincon, 2018). This phenomenon can be observed as light travelling near cosmic massive objects such as black holes, white dwarfs or neutron stars becomes redshifted due to the energy loss within the gravitational field.

Photometric redshifts (photo-z) estimate the redshift of galaxies by analyzing the brightness of galaxies measured through several wide photometric filters or bands. More distant galaxies appear fainter and more redshifted, so by comparing the brightness of a galaxy across different photometric bands, the shift in the galaxy's light can be estimated. This yields the approximate redshift and distance without needing to take detailed spectra of each galaxy. Photometric redshifts rely on the characteristic spectral energy distributions of galaxies and how they are redshifted by cosmic expansion. Machine learning techniques can also be used to train models that estimate redshifts from photometric data.

Utilizing Deep learning algorithms using the spectral frequency bands of a galaxy have been explored as a method for distance estimation through redshift prediction. An earlier approach was made by Firth, Lahav, and Somerville in their research work with a goal to estimate photometric redshifts of galaxies using artificial neural networks (ANNs) on a simulated galaxy data spanning $0<z<3.5$. Here inclusion of derived features like morphological parameters and band overlaps, along with magnitudes improved the redshift prediction with RMSE of 0.021 as compared to 0.023 obtained without any morphology data (Firth, Lahav, and Somerville, 2003). These outcomes demonstrate the effectiveness of ANNs for photometric redshift analysis, matching or exceeding standard template-fitting methods.

Some recent works in this field includes the development of a machine learning framework using ExtraTreesRegressor machine learning algorithm for accurately estimating photometric redshifts of galaxies spanning a wide range from low to very high redshifts ($0 < z < 7$; Reza and Haque, 2020). This yielded root mean squared error (RMSE) of 0.81 on the model validation set on a framework that demonstrated reliable photometric redshift inference even for high redshifts, despite limited spectroscopic training data.

With the above results and promising potential of further accuracy and precision gains associated with the introduction of further morphological parameters to the model, this project work aims to fuse the redshift prediction ANN model utilizing varying input sets based on flux information parameters with a deep learning architecture utilizing image stamps of galaxies to learn and better predict redshift data which infers on better distance estimation. The distance can be calculated given the redshift values using the Hubble's law formula where:

$$D \approx (C*Z) / H\_0 \quad (1)$$

Where:

- D is the estimated distance to the object.
- c is the speed of light in vacuum (approximately 3 x 10^5 km/s).
- z is the redshift of the object.
- H_0 is the Hubble constant, representing the present-day rate of the universe's expansion.

Chapter 3: Methodology

3.1 Data Collection:

The foundational step of this research was to obtain a comprehensive redshift predictor dataset which was attained utilizing the Galaxy and Mass Assembly (GAMA) data. GAMA is a project to exploit the latest generation of ground-based and space-borne survey facilities to study cosmology and galaxy formation and evolution (GAMA Collaboration, 2023). The data is a multispectral imaging and spectroscopic redshift survey in the FUV, NUV, U, G, R, I, Z, X, Y, J, H, K, W1,W2,W3,W4,P100,P160,S250,S350 and S500 bands. The data was categorized using a selection criteria r = 8.9 - 2.5*log10(flux_rt) > 19.5 and uberclass = galaxy, star OR ambiguous where the r < 20.5 cut this was chosen because this is the limit to which the optical spectrum band photometry has been extracted. The survey data also contains the corresponding uncertainties of all measure frequency flux band of each GAMA object in this catalogue correspond to the DFAErr values in the individual filter catalogues as well as the GAMA objects positional variables indicating its right ascension and declination in the celestial sphere. There are precisely 2,232,985 data entries in the dataset comprising of 2232985 rows × 65 columns.

Image stamps of the galaxies utilized for this project is obtained from Galaxy And Mass Assembly (GAMA) survey stamps courtesy of Prof. Simon Driver. The criteria used for its selection to pick out these galaxies was r = 8.9 - 2.5*log10(flux_rt) > 19.5 with uberclass =”galaxy”, mask=0 and duplicate==0 which helps selects objects with uberclass labelled galaxy which is brighter than a magnitude threshold of 19.65. This selection comprises of

166,177, image stamps providing sufficient count of morphological parameters for a deep learning model to make a redshift estimate prediction.

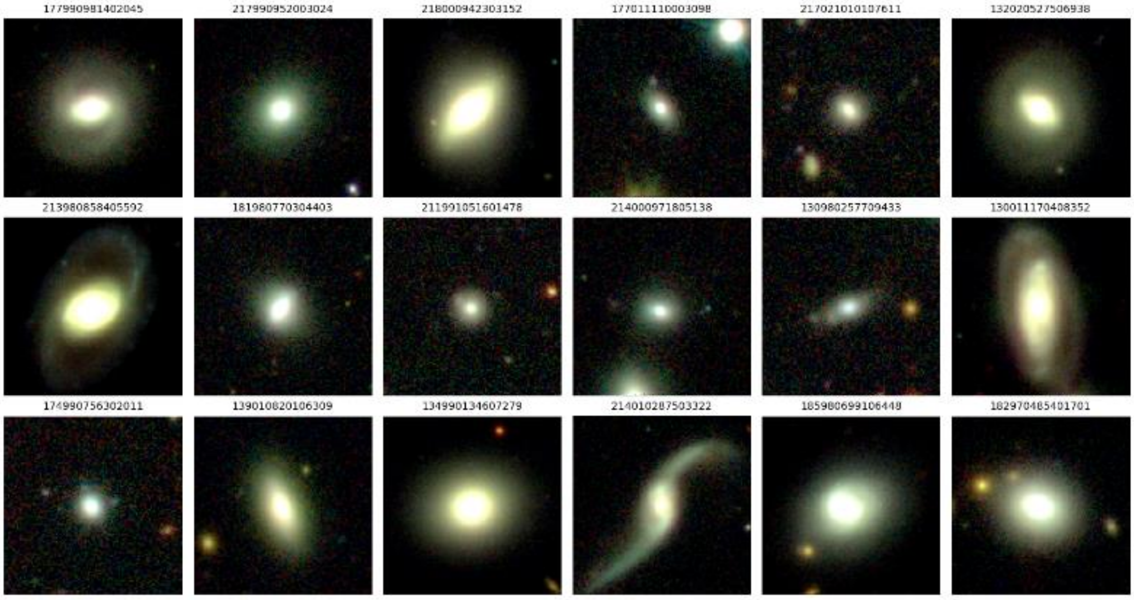


Fig 1: A grid display of a sample of 18 select GAMA image stamps.

3.2 Data Cleaning and Preprocessing

On the input GAMA data utilized for the research work, duplicate values on the CATAID which refers to a unique catalogue identifier number assigned to the matched GAMA object was found. This indicated two or more redshift measurements carried out on the same GAMA object with distinct varying flux and heliocentric redshift readings.

53,604 duplicate GAMA objects were found, to handle duplicate entries in the GAMA catalog, I selected the entry for each duplicate CATAID that minimized the VEL_ERR attribute and maximized the "estimated probability of redshift success", keeping only the top performing entry for each CATAID and dropping the lower outcome duplicates.

The above was implemented utilizing a new attribute to each duplicated Gama object called “compute outcome” using the formulae:

$$(2) \qquad Compute\_Outcome = PROB - (VEL\_ERR / 100)$$

Where:

- PROB = Estimated probability of redshift success
- VEL_ERR = Uncertainty in redshift, given as a velocity, delta c ln(1+z)
- The VEL_ERR maximum value was 119.2 but was clipped at 100 to ensure for a unity (1) maximum outcome.

- the COMPUTE_OUTCOME subtracts a scaled version of VEL_ERR from PROB. The scaling brings VEL_ERR into a 0-1 range by dividing by 100.

This creates a metric that favors rows with high PROB (good redshift estimation) and low VEL_ERR (low velocity error). This was then sorted by COMPUTE_OUTCOME in descending order so the best rows according to this metric are first, and subsequent values are dropped.

In the Survey data catalogue, objects that have no measurement in a particular band are marked with values of -999 in the flux band and its corresponding uncertainty column. To resolve this, K-Nearest Neighbors (KNN) algorithm was utilized on all flux density band columns. With the desired number of neighbors K set to 10 and the GAMA data sorted based on the next adjacent frequency band to ensure that a positive correlation is attained owing to minimized shift in the measure frequency spectrum which ensures that the GAMA Objects within nearby samples within the data are identical thereby improving the accuracy and reliability of the results obtained utilizing the KNN algorithm.

The uncertainties linked to flux density measurements across various frequency bands in the GAMA dataset as shown in fig 2, represent the inherent imprecision in estimating the flux density of GAMA objects at those specific frequency bands. Focusing on the error margins, we directed our attention to data points exceeding the third quartile, which exhibited elevated error rates. To impute flux values for measurements within these large errors, we leveraged the K-Nearest Neighbors (KNN) algorithm with a parameter KK set to 10 to predict the corresponding flux band values.

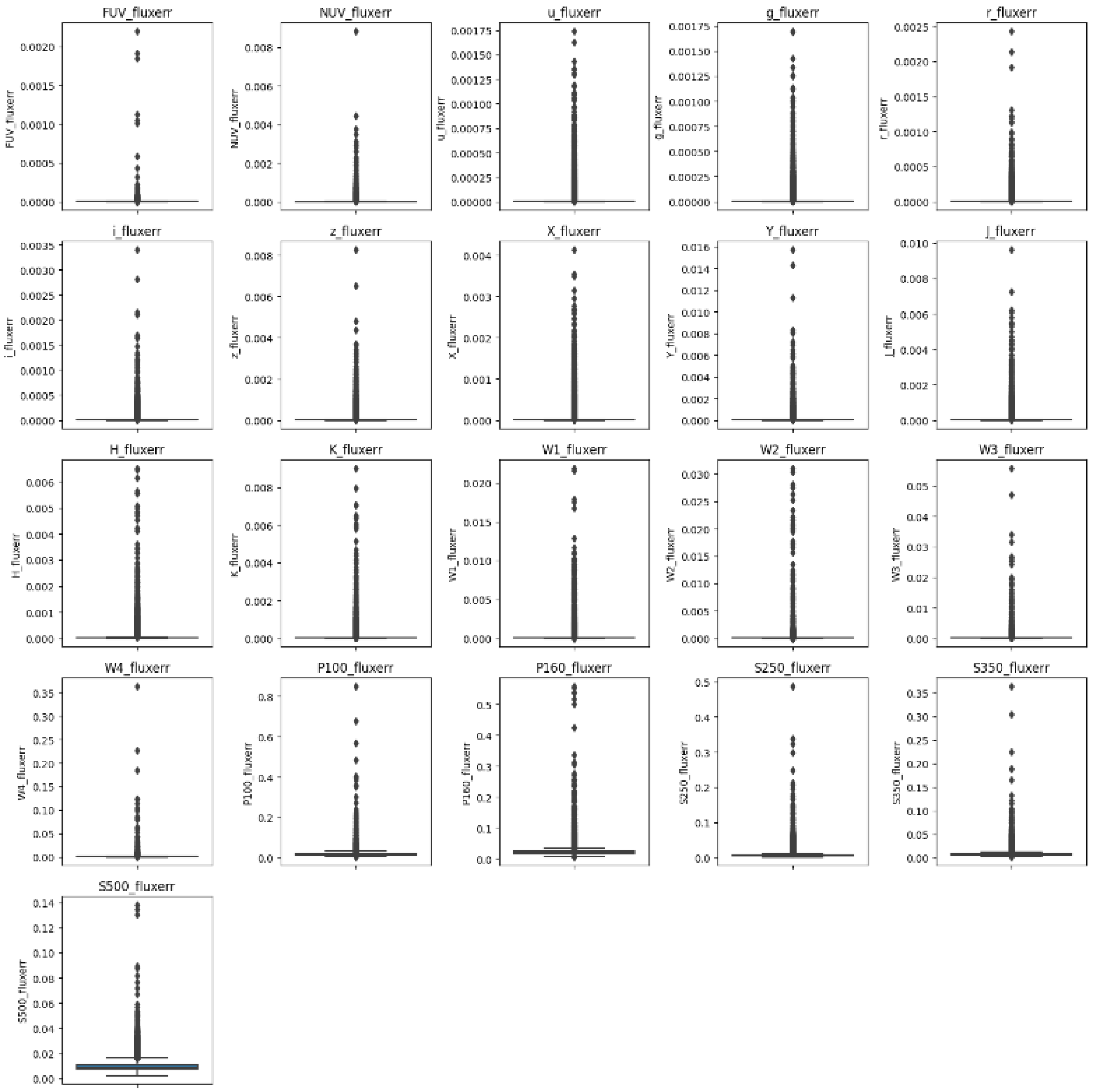


FIG 2: A grid of subplots containing the boxplot representing the distribution of values for the flux error and uncertainty on each frequency band.

3.3 Exploratory Data Analysis (EDA)

The initial stage of our scientific investigation involved a thorough examination of the connections between the flux bands employed for predicting redshifts in the GAMA flux catalog. This exploration was undertaken and represented using a confusion matrix.

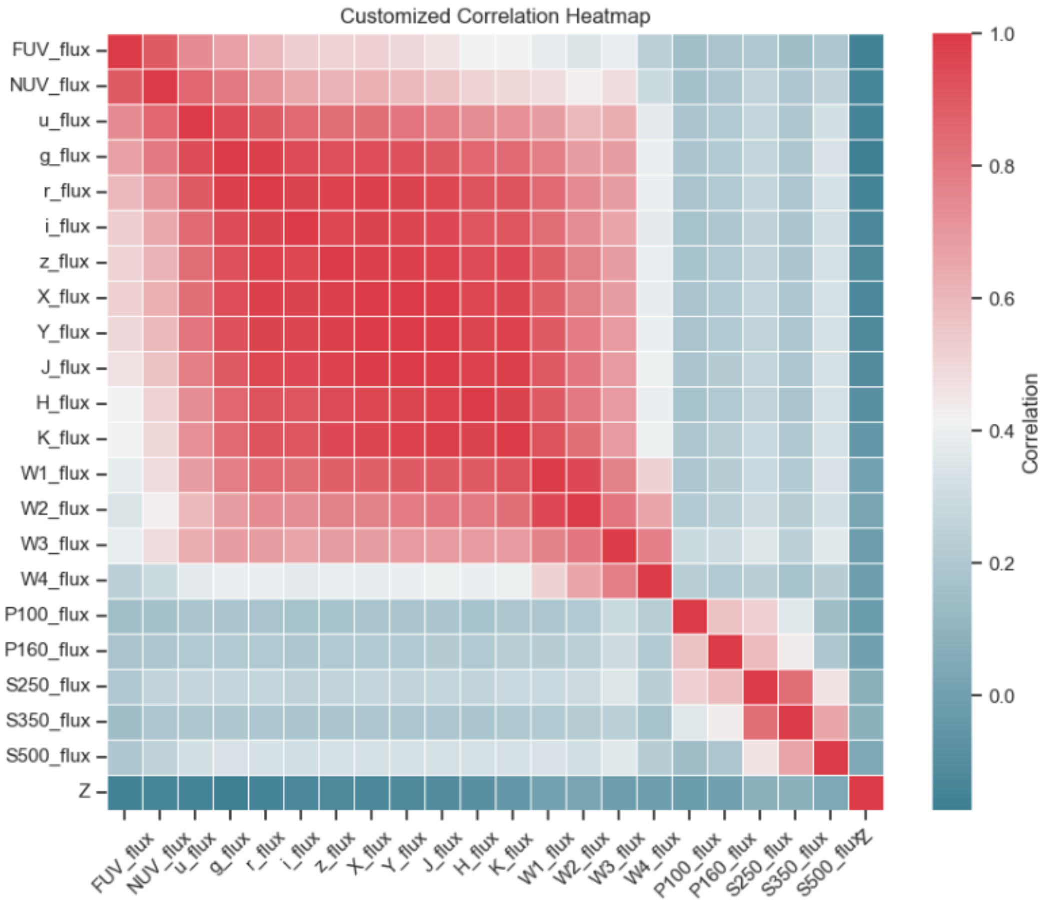


Fig 3: Heatmap showcasing the correlation of the band fluxes within the GAMA data

The analysis of the data from the sample GAMA (Galaxy and Mass Assembly) objects has revealed a distinctive correlation cluster among the high-frequency, low-wavelength flux bands. This cluster indicates that flux bands with range from 0.01 nm to 22 µm bands are closely related, exhibiting strong positive correlations. Such correlations, when present in regression or modeling tasks, can give rise to multicollinearity—an issue where predictor variables are highly interrelated. This can introduce undesirable effects, including overfitting, reduced interpretability, and inflated standard errors of regression coefficients.

To address this multicollinearity challenge and extract meaningful information from the correlated flux bands, we employed Principal Component Analysis (PCA), a powerful dimensionality reduction technique to the GAMA data.

3.4 The proposed model

The model implementation is categorized under three major phases which is inclusive of the following:

- **GAMA flux band Redshift 'Z' Predictor**: The preprocessed GAMA flux density data for each galaxy was used to train machine learning models to predict the redshift values. Multiple algorithms were explored, including random forest regressors, support vector machines, and artificial neural networks. Hyperparameter tuning was conducted to optimize model performance. The Random Forest demonstrated the best results, with low bias and variance. It captured nonlinear relationships between flux densities and redshift effectively to

- **The image 'Z' predictor model**: The galaxy image stamps were fed into a deep convolutional neural network regressor to extract morphological features correlated with redshift. Data augmentation was used to expand the training set. The CNN automatically learned hierarchical image features relevant for redshift prediction through its layered architecture. Spatial context and texture patterns associated with galaxy evolution were encoded. Key imaging augmentation techniques included rotations, flips, zooms, and translations. Batch normalization and dropout were implemented to optimize convergence and generalization.

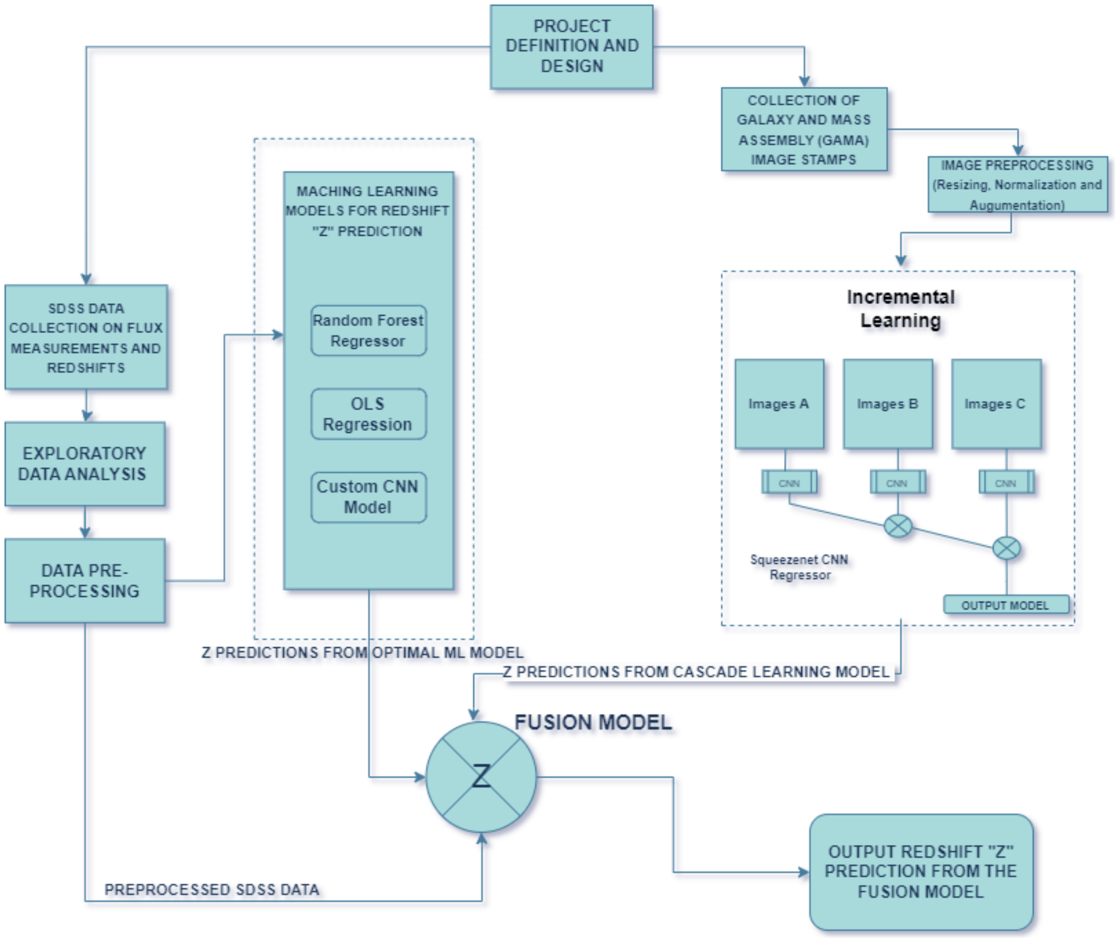

PROJECT DEFINITION AND DESIGN
COLLECTION OF GALAXY AND MASS ASSEMBLY (GAMA) IMAGE STAMPS
IMAGE PREPROCESSING (Resizing, Normalization and Augumentation)
MACHING LEARNING MODELS FOR REDSHIFT "Z" PREDICTION
Random Forest Regressor
OLS Regression
Custom CNN Model
SDSS DATA COLLECTION ON FLUX MEASUREMENTS AND REDSHIFTS
EXPLORATORY DATA ANALYSIS
DATA PRE-PROCESSING
Incremental Learning
Images A
Images B
Images C
CNN
CNN
CNN
Squeezenet CNN Regressor
OUTPUT MODEL
Z PREDICTIONS FROM OPTIMAL ML MODEL
Z PREDICTIONS FROM CASCADE LEARNING MODEL
FUSION MODEL
Z
PREPROCESSED SDSS DATA
OUTPUT REDSHIFT "Z" PREDICTION FROM THE FUSION MODEL

Fig 4: The multimodal Machine Learning Pipeline

- **The fusion model**: The flux band ANN regressor and image CNN regressor outputs were finally integrated using a fusion architecture. This enabled leveraging the complementary strengths of both photometric data and visual morphology. The fusion model weights and combines the predicted probabilities from the two sub-models. It was trained end-to-end, further refining the combined predictions. This multi-modal approach aimed to improve robustness, accuracy and reliability over single-modality models.

The three-phase design provides a systematic methodology for exploiting both galaxy survey photometry and image data to optimize cosmological redshift estimation.

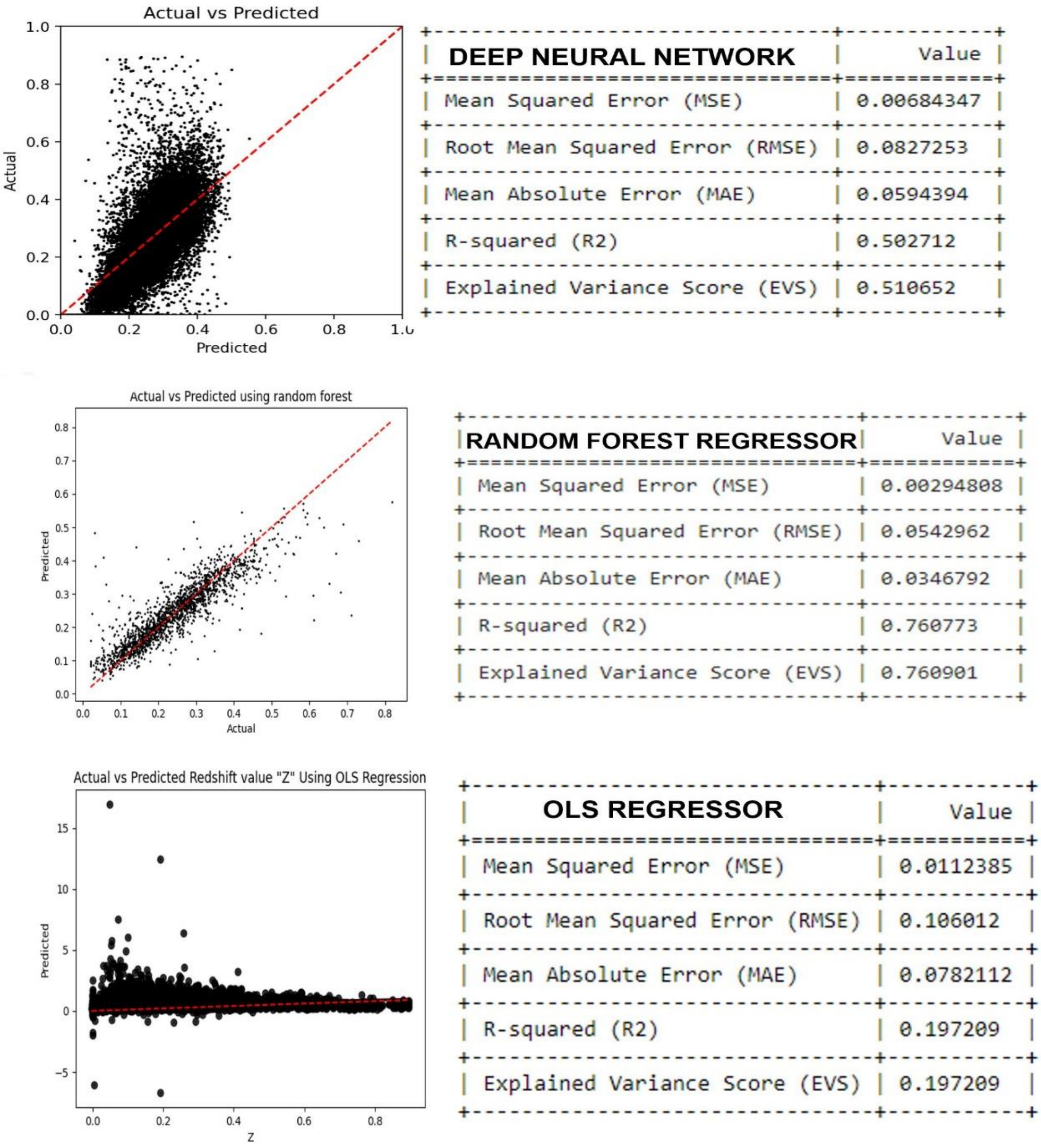
3.4.1 GAMA flux band Redshift 'Z' Predictor:

Three models were implemented at this stage to compare output for the best redshift predictor utilizing the flux band density GAMA data. This includes. A deep neural CNN model, an OLS regression and a Random Forest regressor.

A deep neural network was modeled this consists of an input layer that takes in a 21-dimensional flux density feature vector for each galaxy. This feeds into two fully connected hidden layers of 64 nodes each, with 'ReLU' activation to introduce nonlinearity. Dropout of 0.2 after the first hidden layer regularizes the model to prevent overfitting. The output layer is a single continuous node with a linear activation function to regress the redshift values. The model is compiled with the 'Adam' optimizer, mean squared error loss, and MSE evaluation metric for training and validating performance. Overall, the model provides a flexible nonlinear function approximator to learn the mapping between galaxy photometry and cosmological redshift.

The second model implements a random forest regressor to predict redshift values from GAMA galaxy flux measurements. The flux densities in 21 optical and infrared bands are set as the feature space. The GAMA redshift serves as the continuous target variable. The data is split 80/20 into train and test sets for modeling. The random forest consists of 100 decision tree estimators trained on bootstrapped samples of the training data. This ensemble method averages the predictions from individual trees to improve generalization and reduce overfitting. Each tree recursively partitions the feature space to maximize information gain at splits. The trained forest model is used to predict redshifts on the held-out test data.

The third model implements ordinary least squares linear regression to predict redshift from 21 GAMA galaxy flux band measurements. The features are standardized to have zero mean and unit variance. A constant term is added to capture the bias. The model is trained on 80% of data and validated on 20% to optimize the linear mapping between fluxes and redshift. By minimizing the residual sum of squares, the optimal regression coefficients are obtained. This provides an interpretable parametric model of the linear relationships between photometric properties and redshift.



```
+---------------------------------+------------+
| DEEP NEURAL NETWORK             |      Value |
+=================================+============+
| Mean Squared Error (MSE)        | 0.00684347 |
+---------------------------------+------------+
| Root Mean Squared Error (RMSE)  | 0.0827253  |
+---------------------------------+------------+
| Mean Absolute Error (MAE)       | 0.0594394  |
+---------------------------------+------------+
| R-squared (R2)                  | 0.502712   |
+---------------------------------+------------+
| Explained Variance Score (EVS)  | 0.510652   |
+---------------------------------+------------+
```

```
+--------------------------------+------------+
|RANDOM FOREST REGRESSOR         |      Value |
+================================+============+
| Mean Squared Error (MSE)       | 0.00294808 |
+--------------------------------+------------+
| Root Mean Squared Error (RMSE) | 0.0542962  |
+--------------------------------+------------+
| Mean Absolute Error (MAE)      | 0.0346792  |
+--------------------------------+------------+
| R-squared (R2)                 | 0.760773   |
+--------------------------------+------------+
| Explained Variance Score (EVS) | 0.760901   |
+--------------------------------+------------+
```

```
+--------------------------------+-----------+
|         OLS REGRESSOR          |     Value |
+================================+===========+
| Mean Squared Error (MSE)       | 0.0112385 |
+--------------------------------+-----------+
| Root Mean Squared Error (RMSE) | 0.106012  |
+--------------------------------+-----------+
| Mean Absolute Error (MAE)      | 0.0782112 |
+--------------------------------+-----------+
| R-squared (R2)                 | 0.197209  |
+--------------------------------+-----------+
| Explained Variance Score (EVS) | 0.197209  |
+--------------------------------+-----------+
```

Fig 5: Accuracy plots and metrics using the Deep neural network, Random Forest regressor and OLS regression approach.

In conclusion, the random forest regressor emerged as the optimal machine learning model for predicting galaxy redshifts from photometric data owing to its superior redshift prediction RMSE of 0.0542 and an R2 score of 0.760 in this study. Across the performance metrics of R-squared, root mean squared error, and mean absolute error on the held-out test data, the random forest consistently outperformed the other approaches within this study. The non-parametric tree-based ensemble method proved highly effective at handling the nonlinear relationships and complex interactions between the 21 GAMA flux band measurements and the redshift target variable. By averaging the predictions from many decentralized decision trees, the random forest avoided overfitting and achieved robust generalization. The trained model could accurately predict redshifts across a diverse dataset spanning a wide range of galaxy types and redshifts. Given its flexibility, nonlinear modeling capabilities, and ensemble-based regularization, random forest regression shows great promise for precision redshift estimation using photometric galaxy surveys.

3.4.2 The image ‘Z’ predictor model

An incremental learning approach was used where training occurred sequentially on three subsets of the GAMA image file, the model will be able to learn and update with every new data provided. (Ade and Deshmukh, 2013). The SqueezeNet CNN architecture was leveraged as the training model, with each successive model leveraging the previous model's learned weights as a pretrained starting point. This allows incrementally enhancing the model's representations and performance through multiple generations of transfer learning.

Each subset GAMA image files were passed into the data generators to perform real-time data augmentation on the training, test and validation images. This helps prevent overfitting and improves the generalization of the image-based machine learning model for redshift prediction.

The training and validation data is prepared for the image-based redshift prediction model by matching image filenames to their corresponding redshift labels. A list of all image filenames in the training directory is obtained which is looped through to extract the numeric ID from the name. This ID is looked up to same matching ID in the redshift catalog to get the target redshift value. The matched filename and redshift is appended into a new dataframe.

The dataframe matches the images to their labels so the data can be fed into the image-based model during training. The model will use the images as input and predict the redshift values to match the 'target' labels for backpropagation.

Metric	**Model A**	**Model B**	**Model C (Output model)**
Mean Squared Error (MSE)	0.0339065	0.0297193	0.0197511
Root Mean Square Error (RMSE)	0.184137	0.172393	0.140538
Mean Absolute Error (MAE)	0.0807192	0.0779095	0.0635469
R-squared (R2)	0.414194	0.456683	0.562247
Explained Variance Score (EVS)	0.427383	0.476012	0.562249

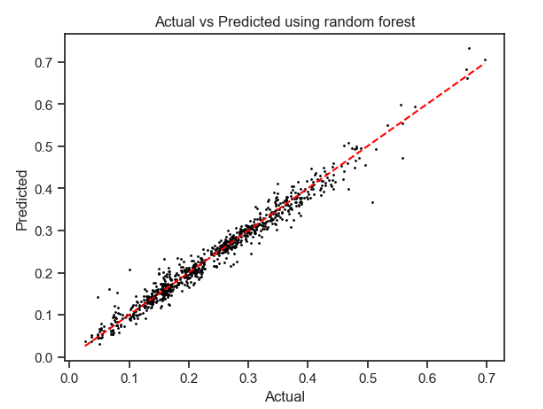
FIG 6: Table displaying the Accuracy metrics for each stage of the incremental learning model approach on GAMA images.

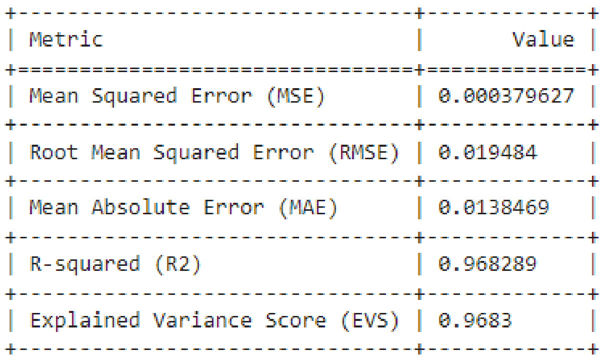
3.4.3 The fusion model

The fusion model integrates predictions from the GAMA flux band model and CNN image model to improve redshift estimation accuracy. The two sets of predicted redshifts are concatenated as features to train a random forest regressor for final redshift prediction. The flux band predictions capture galaxy color and photometric properties. The image model contributes morphological features like spiral arms and elliptical shapes. Together they provide complementary information.

The trained fusion model demonstrates improved performance over individual models, validating the benefit of multi-modal data integration. The non-parametric flexibility of random forest handles complex feature interactions.

In summary, the fusion model harnesses diverse predictive signals from both galaxy images and photometry for robust redshift inference. The random forest effectively integrates these heterogeneous data sources.





```
+--------------------------------+-------------+
| Metric                         |       Value |
+================================+=============+
| Mean Squared Error (MSE)       | 0.000379627 |
+--------------------------------+-------------+
| Root Mean Squared Error (RMSE) | 0.019484    |
+--------------------------------+-------------+
| Mean Absolute Error (MAE)      | 0.0138469   |
+--------------------------------+-------------+
| R-squared (R2)                 | 0.968289    |
+--------------------------------+-------------+
| Explained Variance Score (EVS) | 0.9683      |
+--------------------------------+-------------+
```

Figure 7: Accuracy plots and metrics using the Fusion Model

Chapter 4: Results and Discussion

The multi-modal model achieved a lower root mean squared error (RMSE) of 0.019484 and higher R2 of 0.9683 on the test set compared to its input models. The image-only CNN had an RMSE of 0.046 and R2 of 0.71, while the flux-only model scored 0.0294808 and 0.7698 on RMSE and R2 respectively as shown in table below.

Metric	Flux band Z predictor model	GAMA image Z predictor model	The fusion model
Mean Squared Error (MSE)	0.00294808	0.0197511	0.0000379627
Root Mean Square Error (RMSE)	0.0542962	0.140538	0.019484
Mean Absolute Error (MAE)	0.0346792	0.0635469	0.0138469
R-squared (R2)	0.760773	0.562247	0.968289
Explained Variance Score (EVS)	0.760901	0.562249	0.9683

FIG 8: Table comparing the Accuracy metrics of Flux band Z predictor model and GAMA image Z predictor model to the resulting fusion model.

The study indicates that multimodal fusion methods could often result in more accurate galaxy distance estimations compared to utilizing a single mode. This is visualized in fig 9, a plot of the linear regressions of the various prediction models as compared to the actual redshift fit.

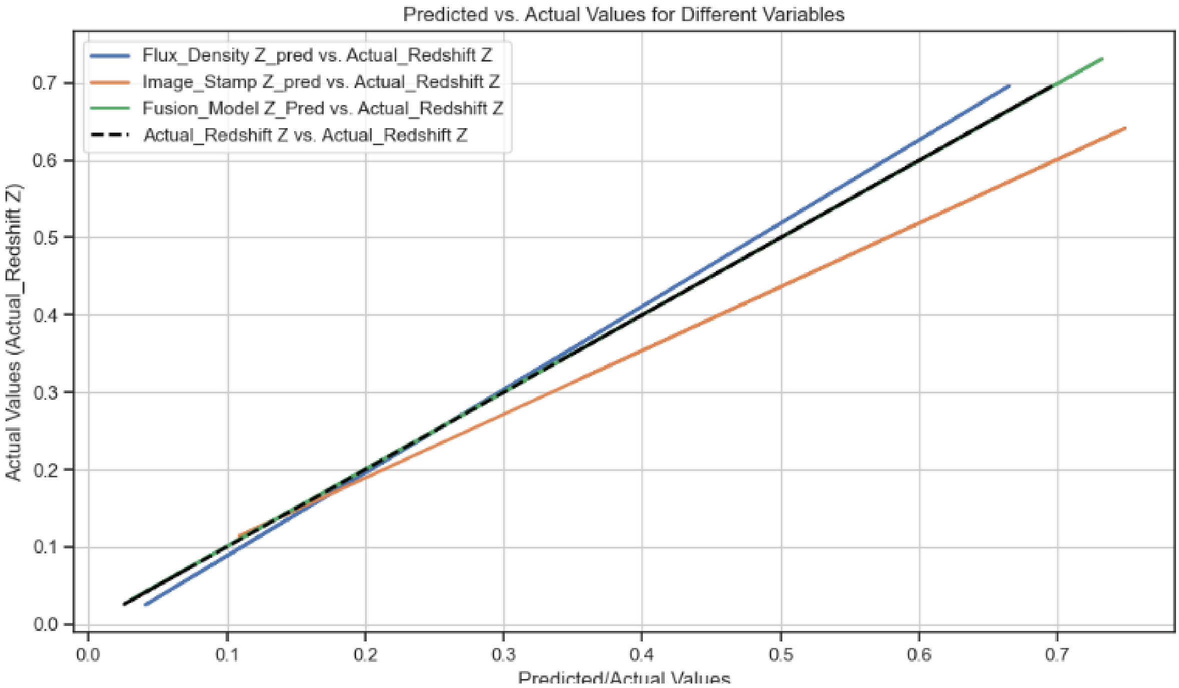


FIG 9: Predicted vs Actual values for different models' variables.

The superior performance of the fusion model highlights the benefits of integrating morphological and photometric galaxy data for redshift prediction. The flux-only model is limited by photometric errors and intrinsic variations in galaxy SEDs. The image model enables extracting morphological indicators related to galaxy evolution, overcoming photometric ambiguities. Fusing the multi-wavelength photometry and visual data provides more predictive signals to constrain redshifts. The random forest regressor effectively integrates these heterogeneous inputs for improved generalization.

Chapter 5: Conclusion and Recommendation

This research demonstrates that a multi-modal machine learning approach can significantly improve the accuracy of galaxy redshift estimation compared to individual models by over 27% as shown in result using the EVS and R2 metrics. The fusion model achieved a superior performance with a 98.71% and 99.81 error (MSE) reduction when compared to the Flux band resulting to a higher correlation to true redshifts.

The study validates that incorporating morphological indicators from galaxy images provides complementary information to photometric measurements. The two data modalities contain predictive signals that reinforce each other when jointly modeled. The non-parametric random forest effectively combined these heterogeneous inputs. While the current framework has limitations in terms of model complexity and dataset size, this work provides a robust proof-of-concept. It highlights the merits of leveraging cross-disciplinary astronomical data in a synergistic fashion to tackle cosmological distance inference problems.

Further research could explore the effects of using velocity distributions that also provide important clues for distance estimates. Galaxy rotational velocities, orbital velocities, and proper motions can complement the flux and morphological information, ultimately aiding to improve prediction accuracy.

REFERENCES:

- Ade, R.R. and Deshmukh, P.R., 2013. Methods for incremental learning: a survey. *International Journal of Data Mining & Knowledge Management Process*, *3*(4), p.119.

- Buongiorno, C. (2023) “Astronomers discover the most distant galaxy yet,” Astronomy Magazine [Preprint]. Available at: https://www.astronomy.com/science/astronomers-discover-the-most-distant-galaxy-yet/. [Accessed 16/08/2023].

- Firth, A.E., Lahav, O. and Somerville, R.S., 2003. Estimating photometric redshifts with artificial neural networks. Monthly Notices of the Royal Astronomical Society, 339(4), pp.1195-1202.

- Freedman, W.L. and Madore, B.F., 2010. The hubble constant. Annual Review of Astronomy and Astrophysics, 48, pp.673-710.

- Hubble, E.P., 1927. Cepheids in spiral nebulae. In Publications of the American Astronomical Society (Vol. 5, pp. 261-264).

- Hubble, E., 1929. A relation between distance and radial velocity among extra-galactic nebulae. Proceedings of the national academy of sciences, 15(3), pp.168-173.

- Ishida, E.E., 2019. Machine learning and the future of supernova cosmology. *Nature Astronomy*, *3*(8), pp.680-682.

- Jarvis, S.H., 2020. Space, and the Redshift Effect.

- Marquez, E.S., Hare, J.S. and Niranjan, M., 2018. Deep cascade learning. IEEE transactions on neural networks and learning systems, 29(11), pp.5475-5485.

- Reza, M. and Haque, M.A., 2020. Photometric redshift estimation using ExtraTreesRegressor: Galaxies and quasars from low to very high redshifts. *Astrophysics and Space Science*, *365*(3), p.50.

- Rincon, B.P. (2018) "Einstein theory passes black hole test," BBC News, 26 July. Available at: https://www.bbc.co.uk/news/science-environment-44967491. [Accessed 01/08/2023].

- "The Doppler Effect in Astronomy" (2022). British Astronomical Association. Available at: https://britastro.org/2022/the-doppler-effect-in-astronomy (Accessed: [07/09/2023]).

- Whiting, A.B., 2004. The expansion of space: Free particle motion and the cosmological redshift. arXiv preprint astro-ph/0404095

- York, D.G., Adelman, J., Anderson Jr, J.E., Anderson, S.F., Annis, J., Bahcall, N.A., Bakken, J.A., Barkhouser, R., Bastian, S., Berman, E. and Boroski, W.N., 2000. The sloan digital sky survey: Technical summary. The Astronomical Journal, 120(3), p.1579.