

Gamma-ray blazar classification using machine learning with advanced weight initialization and self-supervised learning techniques

Gopal Bhatta ¹★, Sarvesh Gharat ², Abhimanyu Borthakur³ and Aman Kumar ⁴

¹Janusz Gil Institute of Astronomy, University of Zielona Góra, ul. Szafrana 2, PL-65-516 Zielona Góra, Poland

²Centre for Machine Intelligence and Data Science, Indian Institute of Technology Bombay, Mumbai 400076, Maharashtra, India

³Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal 576104, Karnataka, India

⁴Department of Electrical Engineering, Tezpur University, Tezpur 784028, Assam, India

Accepted 2024 January 3. Received 2023 December 3; in original form 2023 September 30

ABSTRACT

Machine learning has emerged as a powerful tool in the field of gamma-ray astrophysics. The algorithms can distinguish between different source types, such as blazars and pulsars, and help uncover new insights into the high-energy universe. The Large Area Telescope onboard the *Fermi* gamma-ray telescope has significantly advanced our understanding of the Universe. The instrument has detected a large number of gamma-ray-emitting sources, among which a significant number of objects have been identified as active galactic nuclei. The sample is primarily composed of blazars; however, more than one-third of these sources are either of an unknown class or lack a definite association with a low-energy counterpart. In this work, we employ multiple machine learning algorithms to classify the sources based on their other physical properties. In particular, we utilized smart initialization techniques and self-supervised learning for classifying blazars into BL Lacertae (BL Lac, also BLL) objects and flat-spectrum radio quasars (FSRQs). The core advantage of the algorithm is its simplicity, usage of minimum number of features and easy deployment due to lesser number of parameters without compromising on the performance along with increase in inference speed (at least seven times more than existing algorithms). As a result, the best-performing model is deployed on multiple platforms so that any user irrespective of their coding background can use the tool. The model predicts that out of the 1115 sources of uncertain type in the 4FGL-DR3 catalogue, 820 can be classified as BL Lacs and 295 can be classified as FSRQs.

Key words: radiation mechanisms: non-thermal – methods: observational – methods: statistical – galaxies: active – BL Lacertae objects: general – quasars: supermassive black holes.

1 INTRODUCTION

Blazars, belonging to the class of active galactic nuclei (AGNs), stand out as some of the most luminous and exceptionally variable sources in the Universe. These sources are recognized for their high luminosity, broad-spectrum emissions, and significant rapid variability across a wide range of the electromagnetic spectrum (see e.g. Bhatta, Mohorian & Bilinsky 2018; Bhatta & Dhital 2020; Bhatta 2021). These exceptional characteristics are frequently associated with the emission boosted by Doppler effects from the relativistic outflows originating near the central engine (Urry & Padovani 1995; Jorstad et al. 2017). Conventionally, these objects are typically divided into two main groups: BL Lacertae (BL Lacs) and flat-spectrum radio quasars (FSRQs).

The primary distinction between these two categories lies in the fact that BL Lacs typically display either no or very faint emission line spectra, whereas FSRQs commonly exhibit broad emission lines

and their synchrotron peak is at lower frequencies. While FSRQs are more powerful sources, BL Lacs belong to an extreme class characterized by an excess of high-energy emissions, ranging from hard X-rays to TeV energies. In leptonic models of blazar, this excess arises from synchrotron and inverse-Compton (IC) processes. Their seemingly low luminosity may be attributed to the absence of strong circumnuclear photon fields and relatively low accretion rates (Ghisellini et al. 2011).

Blazars are known for emitting non-thermal radiation across a wide spectrum, spanning from radio waves to TeV gamma-rays. The spectral energy distributions (SEDs) of blazars exhibit two distinct peaks. The first peak, observed in the infrared to soft X-ray energy range, originates from synchrotron emission. In contrast, the second peak, situated in the hard X-ray to gamma-ray region, is associated with IC radiation, as per the leptonic model (see Böttcher 2019, for a recent review). The photons involved in IC scattering can either arise from the same group of electrons responsible for generating the synchrotron peak, as explained by the synchrotron self-Compton (SSC) model (Maraschi, Ghisellini & Celotti 1992; Bloom & Marscher 1996), or they can originate from external sources, including, but not limited to, the accretion disc (Dermer &

* E-mail: g.bhatta@ia.uz.zgora.pl

Schlickeiser 1993), the broad-line region (Sikora, Begelman & Rees 1994), or the dust torus (Blażejowski et al. 2000).

An alternative approach to categorizing blazars involves examining their SEDs and considering the synchrotron peak frequency (ν_s). Following this approach, blazars can be classified into three categories: high-synchrotron peaked (HSP; $\nu_s > 10^{15}$ Hz), intermediate-synchrotron peaked (ISP; $10^{14} < \nu_s < 10^{15}$ Hz), or low-synchrotron peaked (LSP; $\nu_s < 10^{14}$ Hz) blazars. It is noteworthy that, in this classification approach, FSRQs primarily fall under the category of LSP. In the unifying scheme known as the blazar sequence, the bolometric luminosity decreases as the sources transition from FSRQs to HSP sources, whereas gamma-ray emissions increase (Fossati et al. 1998; Ghisellini et al. 2017).

The *Fermi* Large Area Telescope (LAT), in operation since 2008, conducts a continuous survey of the entire sky, identifying gamma-ray sources within the energy range spanning from tens of MeV to the TeV range (Atwood et al. 2009). Among the extragalactic gamma-ray sources observed by the *Fermi* gamma-ray space telescope, blazars constitute the largest population (Abdollahi et al. 2020). A significant number, roughly one-third, of the blazar candidates identified by *Fermi*-LAT up to this point belong to uncertain category. Blazar candidates of uncertain type (BCU) are blazar candidates that do not clearly fit into one of the established blazar subtypes, such as BL Lac or FSRQ. In other words, these blazar candidates may have characteristics that make it difficult to definitively classify them into a specific blazar subtype. One often needs additional data and analysis to determine the precise nature of these objects. In some cases, this might prove to be an expensive and challenging task, especially when dealing with sources that could potentially be of the BL Lac type, which inherently have weak or no emission lines.

As an alternative approach, several authors have recently turned to various machine learning (ML) algorithms to classify BCUs into BL Lacs and FSRQs. For example, Cooper et al. (2023) used the Multivariate Imputation by Chain Equations (MICE) and k -nearest neighbours algorithms to initially fill in missing variables, such as redshift and the highest energy, and subsequently classified AGNs into either BL Lacs or FSRQs using multiple algorithms based on the SuperLearner. In another study, Sahakyan, Vardanyan & Khachatryan (2023) employed Artificial Neural Networks (ANNs), Extreme Gradient Boosting (XGBOOST), and Light Gradient-Boosting Machine (LIGHTGBM) algorithms to classify BCUs into 825 BL Lac candidates and 405 FSRQ candidates, along with 190 sources that remained without a clear prediction. Agarwal (2023) used multiple supervised ML algorithms and classified a sample of 1115 BCUs into 610 BL Lac objects and 333 FSRQs. Additionally, Butter et al. (2022) used Bayesian neural networks for the classification of *Fermi*-LAT blazar candidates into BL Lacs and FSRQs while also estimating associated uncertainties. In a separate study, Kovačević et al. (2020) employed a supervised ML method based on an ANN on a sample of 1329 BCUs, predicting that 801 sources are BL Lacs, 406 sources are FSRQs, and 122 sources remain unclassified. Finally, Kang et al. (2019) employed supervised ML algorithms to generate predictive models, classifying BCUs into 724 BL Lac-type candidates, 332 FSRQ-type candidates, and 256 without a clear prediction.

In this study, we use data from the fourth *Fermi* catalogue (4FGL; Abdollahi et al. 2020) and employ ML-based classification methods to distinguish between BCUs, classifying them as either BL Lacs or FSRQs. The method majorly depends on using of smarter ways of initializing weights and employing self-supervised learning. It has been observed that from all the employed techniques, bias initialization with soft voting happens to perform the best while giving the

accuracy of 93 per cent and macro average $F1$ score of 0.914. Other methods deliver a similar performance with accuracy ranging from 88.6 per cent to 91.5 per cent. The core advantage of our method, in comparison to other available methods, lies in its simplicity, facilitating easy deployment and acceleration of the inference speed. Additionally, our approach emphasizes reproducibility, enabling the replication of results, and allows for direct adoption in scenarios involving bulk predictions. By utilizing a minimal set of features, our method ensures the incorporation of all available sources from the catalogue without compromise.

In this paper, we organize our content as follows: Section 2 introduces the data set and outlines our classification methods, including both vanilla architectures and intelligent modifications, as well as our training and testing strategies. Section 3 presents our results and discusses their implications. Finally, in Section 4 we outline our conclusions.

2 METHODOLOGY

2.1 Data collection and processing

In this study, we make use of *Fermi*'s fourth catalogue of AGN data release 3 (4LAC-DR3;¹ Ajello et al. 2022), which is based on data accumulated over 12 yr and contains over 6600 sources. The catalogue comprises 1458 BL Lac objects, 792 FSRQs, and 1493 BCUs, which are shown in the sky map in Fig. 1. The catalogue consists of 3407 sources, each defined by a total of 41 observational features. However, 35 features are provided without any missing values. The feature ‘SED_class’ has the highest number of missing values (989), while the ‘Counterpart_Catalog’ feature has the least, with only 20 missing values. However, for this study we only make use of clean samples [refer to Abdollahi et al. (2020) for more information], hence reducing the sample size to 1335 BL Lac objects, 670 FSRQs, and 1115 BCUs, respectively. Further, we observe that although considering all the features can result in better results, it minimizes the number of samples for which we can make predictions due to the missing values, and imputation of it can result in artificial bias making things erroneous. Hence, along with referring to important contributing features in Agarwal (2023) and the availability of data for different features, we proceed with seven features, namely ‘PL_Index, nu_syn, LP_Index, Pivot_Energy, Frac_Variability, Variability_Index, and nuFnu_syn’.² Next, the collected data are divided into train, test, and validation sets in the ratio of 80:10:10. To have a fair evaluation of the proposed methods, we ensure that the test data consist of data that have never been seen by the algorithm previously. Further, all features except PL_Index, LP_Index, and Frac_Variability underwent log transformations, which ensures that the large values of parameters do not explicitly distort the learning of the model.

2.2 Model architecture

The study broadly explores two different approaches: traditional ANNs and self-supervised learning. The former is divided into three parts, primarily focused on various weight initialization strategies,

¹<https://fermi.gsfc.nasa.gov/ssc/data/access/lat/4LACDR3/>

²A description of the observed features is available in table A1 of Ajello et al. (2022), providing detailed information on the characteristics under consideration.

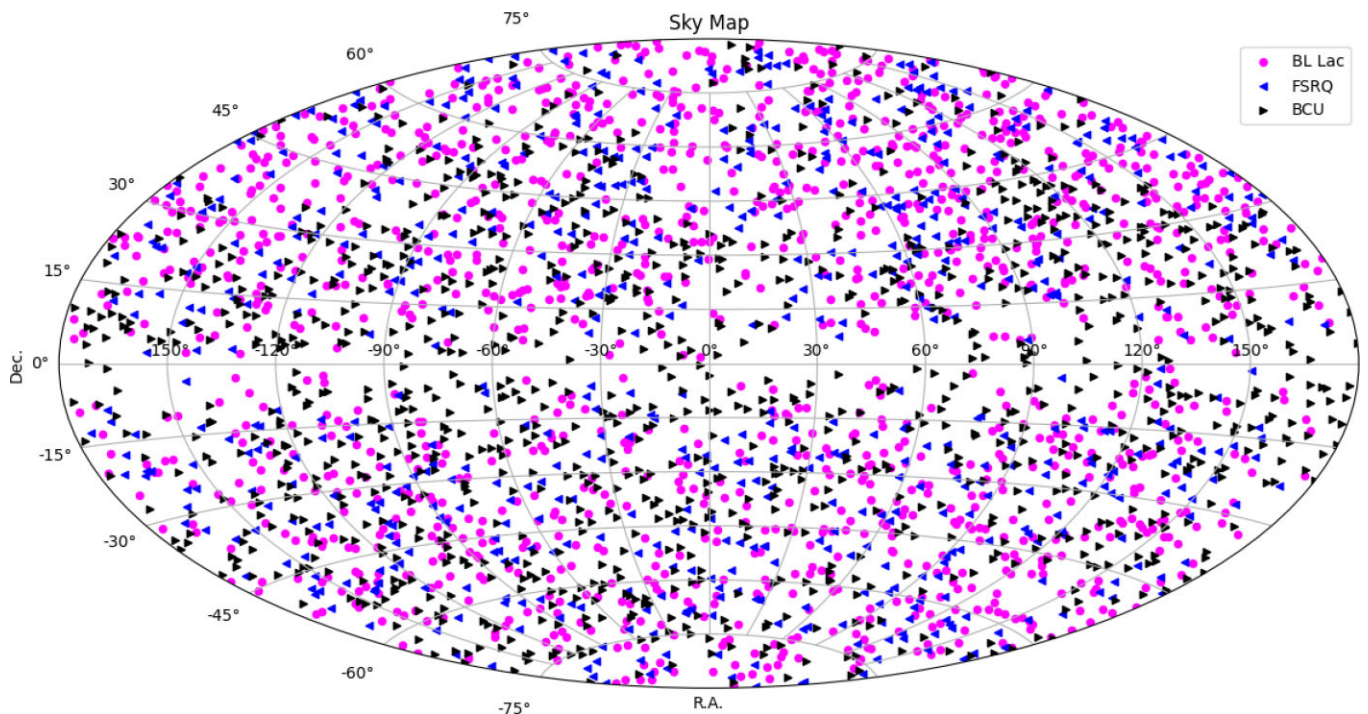


Figure 1. Sky map showing population of BL Lacs, FSRQs, and BCUs from *Fermi*/LAT recent catalogue of AGN 4FGL LAC-DR3.

and is elaborated as follows. The first and the best-performing architecture uses a bias-initialization technique to initialize the weights in the final layer, a common technique for weight initialization when dealing with imbalanced data sets. Since the proposed architecture comprises only 1 hidden layer with 42 neurons, it has a significantly lower number of parameters, making it computationally efficient. In addition to the initialization, we also introduce an ensemble-based soft voting approach for this architecture – after extensive experimentation and saving the training weights at each epoch, we retrieve the architecture’s weights from the 40th and 41st epochs to create two distinct pseudo models. We then evaluate both models on the test data, assigning a weight of 0.1 to the predictions from the model corresponding to the 40th epoch and a weight of 0.9 to the predictions from the 41st epoch. The final prediction is calculated as the weighted sum of these predictions, rounded to the nearest whole number. These weight values are hyperparameters tuned through a trial-and-error method.

The remaining two initialization techniques are applied to an ANN with 2 hidden layers containing 64 and 32 neurons, respectively, with a dropout (Srivastava et al. 2014) of 0.5 in between them. Making use of dropout in the study, ensures that at any given point while training a particular neuron will remain inactive with probability of 0.5, hence, ensuring that there is not any overdependence on a particular neuron thereby encouraging the generalizability of the model. The second initialization technique is a greedy-based pretraining approach in a supervised fashion. Here, initially we train the input layer without considering the hidden layer. This provides us with an estimate of the weights that would be best if there were no hidden layer. Next, while keeping the input layer weights constant, we train the algorithm again to estimate the weights of the neurons in the hidden layer.

The third and final approach for initialization is again a greedy-based approach; however, in this case we deal in an unsupervised manner (Erhan et al. 2010). This is a widely used technique when dealing with data having a large number of unlabelled data points.

While its purpose is to learn the data distribution for weight initialization, there is no guarantee that this method will achieve optimal performance, even when compared to its vanilla version. We observe a similar behaviour when using an autoencoder-based method to pretrain the network in an unsupervised fashion.

In summary, the first bias-based initialization technique is implemented on a traditional ANN having 1 hidden layer with the model having 42 neurons with a dropout of 0.5 and the other two greedy-based approaches are applied to an ANN with 2 hidden layers containing 64 and 32 neurons, respectively, with a dropout of 0.5 in between them. As this is a binary classification task, the loss function used for this model is ‘binary crossentropy’ with the activation function being ‘sigmoid’ in the output layer. Using of sigmoid as an activation function restricts the output between 0 and 1 allowing us to use a single neuron in the output layer. In our specific case, we represent ‘BL Lac’ by ‘0’ and ‘FSRQs’ by ‘1’. Hence, the more the output is towards 1, the more confident is the algorithm in predicting the corresponding input target as FSRQ, similar is the case with BL Lac wherein the output has to be very close to 0.

The second major approach in this study is self-supervised learning, where we employ pretext tasks such as autoencoders (Chen et al. 2023) and contrastive classification (van den Oord, Li & Vinyals 2018; Henaff 2020; Liu et al. 2023). Though autoencoder is an unsupervised algorithm, it has been widely used in a self-supervised regime wherein the autoencoder is used to train the encoder to have a meaningful representation of the data. This is known as the pretext task which is not the primary task the model is aimed at, but helps in generating a better understanding of the data for the downstream tasks. Here, the encoder’s learned representations are fed as an input to a classifier for the main downstream task of classification. For our case, the encoder + decoder architecture consists of seven layers with input and output layers having seven neurons, respectively. Next, the remaining layers of the encoder consist of 128, 64, and 32 neurons, respectively. On the other hand, the decoder has a total of 2 layers prior to output with 64 and 128 neurons, respectively. As



Figure 2. Confusion matrix on testing data for all the models with X axes denoting the predicted label and Y axes denoting the true label.

Table 1. Performance summary: a comparison table between class wise Precision, Recall, *F1* score, Accuracy, and AUC for all the models employed in this study.

Model	Accuracy	BL Lac			FSRQ			AUC
		Precision	Recall	<i>F1</i> score	Precision	Recall	<i>F1</i> score	
Bias initialization with soft voting	0.930	0.938	0.965	0.951	0.909	0.847	0.877	0.949
Supervised greedy pretraining	0.915	0.963	0.915	0.939	0.818	0.915	0.864	0.956
Unsupervised greedy pretraining	0.886	0.941	0.894	0.917	0.773	0.864	0.816	0.956
SSL autoencoder pretext	0.891	0.941	0.901	0.921	0.785	0.864	0.823	0.949
SSL contrastive classification	0.886	0.941	0.894	0.917	0.773	0.864	0.816	0.951

this acts as a regression-based task wherein we need to reconstruct the input the loss function used here is the ‘mean squared error’. Further, the encoded representation having 32 neurons is directly connected to the output layer making use of ‘softmax’ as an activation function.

The final model that we discuss in this study is ‘contrastive classification’. Broadly this method leverages the concept of similarity and dissimilarity between data samples. It aims to bring similar samples closer in the feature space while pushing dissimilar samples apart. For this particular case of classification, we create positive and negative pairs of FSRQs and BL Lacs, where a pair of two FSRQs or two BL Lacs is considered positive pair and a pair of two dissimilar objects i.e. one FSRQ and one BL Lac is considered negative pair. The model then tries to learn how to distinguish between two types of pairs as the pretext task and while doing so it learns about the data which in turn can be used in downstream task. To do so, we create positive and negative pairs and then move ahead with creation of model, for that we define two input layers having the number of neurons equal to the feature size. Both input data samples are then passed through a shared dense layer with 64 units and ReLU activation. Further, the embeddings of the two samples are concatenated and a dense layer with a sigmoid activation is used to predict if the pair is positive or negative. As we now have the features extracted by the contrastive learning model, we then feed it to our standard classifier having connected the extracted features directly to the output layer.

In summary, we propose five different models – three involving ANNs and two which use self-supervised learning. The final class prediction from each model is obtained by applying a threshold of 0.5 on the predicted probability, where all values above 0.5 are classified as FSRQs(1) and those below are classified as BLLs(0).

2.3 Training and validation

The proposed algorithms are implemented using TensorFlow.³ To ensure the reproducibility of the achieved results, we fix a random seed, hence irrespective of the number of times the algorithm is implemented, the trained weights will remain the same. Next, to avoid some over complication, we consider all the data points in a batch while training as opposed to the standard mini-batch approach to calculate the loss and accordingly optimize the model. One of the major reasons behind having done this is the data set being heavily imbalanced, considering a smaller batch size may result in all the samples being of one single class and hence distorting the learning process. Though this can be avoided by doing a batch normalization or by smartly dealing with the sampling process, we prefer to go with a simple method resulting in minimization of computational requirements. Next, to optimize the algorithm we make use of

‘Adam’ (Bae, Ryu & Shin 2019; Mehta, Paunwala & Vaidya 2019) as an optimizer which is one of the majorly used optimizers in the ML community due to obvious reasons. Note that, we do not discuss the reasons in this study as there’s a large literature available on the optimizer commenting about its features like incorporating momentum along with being a variant of Adagrad, which ultimately helps in quicker convergence.

One of the core advantages of our algorithm is the number of parameters which is significantly less when compared to other existing models making our model faster and easier to deploy without compensating for the results. To account for rough estimate on the speed of our algorithm relative to the other algorithms, we calculate the number of floating point operations (FLOPs). FLOPs are generally calculated based on the number of parameters in a neural network model. Having fewer FLOPs is particularly advantageous for accelerating the predictions, making it easy and efficient to use in bulk. The number of FLOPs done while predicting the results is 672, which is significantly lower than Agarwal (2023) (5056). As a result, the best-performing model is further deployed on Streamlit⁴ and Amazon AWS⁵ so that a user can enter the values for all the features and get a prediction along with the prediction probability for the corresponding input.

Next, while training, all the models are trained for a maximum of 1500 epochs with a stopping condition on validation accuracy, ensuring that the model will stop its training if there’s no increase in validation accuracy for 300 epochs. This introduction of stopping criteria, along with the implemented dropout of 0.5 helps the algorithm avoid overfitting and also avoid unnecessary computations which would not improve the results.

3 RESULTS AND DISCUSSION

The proposed method is evaluated on the test sample that has never been seen by the algorithms during its training. One of the major reasons to have a completely independent test data is to avoid any biases that would occur if the model had updated its weights on the same data. The test data consist of 142 BL Lacs and 59 FSRQs that approximately comprise of 10 per cent of the labelled data. The test data are randomly chosen from the labelled data such that it contains all the possible input distributions in a generalized fashion.

For each model, we present a corresponding confusion matrix shown in Fig. 2. As seen in the figure, the bias initialization with soft voting corresponds to lowest misclassification rate. Though, the number of FSRQs predicted correctly by the algorithm is the lowest compared to other models, the numbers just differ by a few samples, and hence this does not prove to be a barrier. Next,

³<https://www.tensorflow.org/>

⁴<https://streamlit.io/>

⁵<https://aws.amazon.com/>

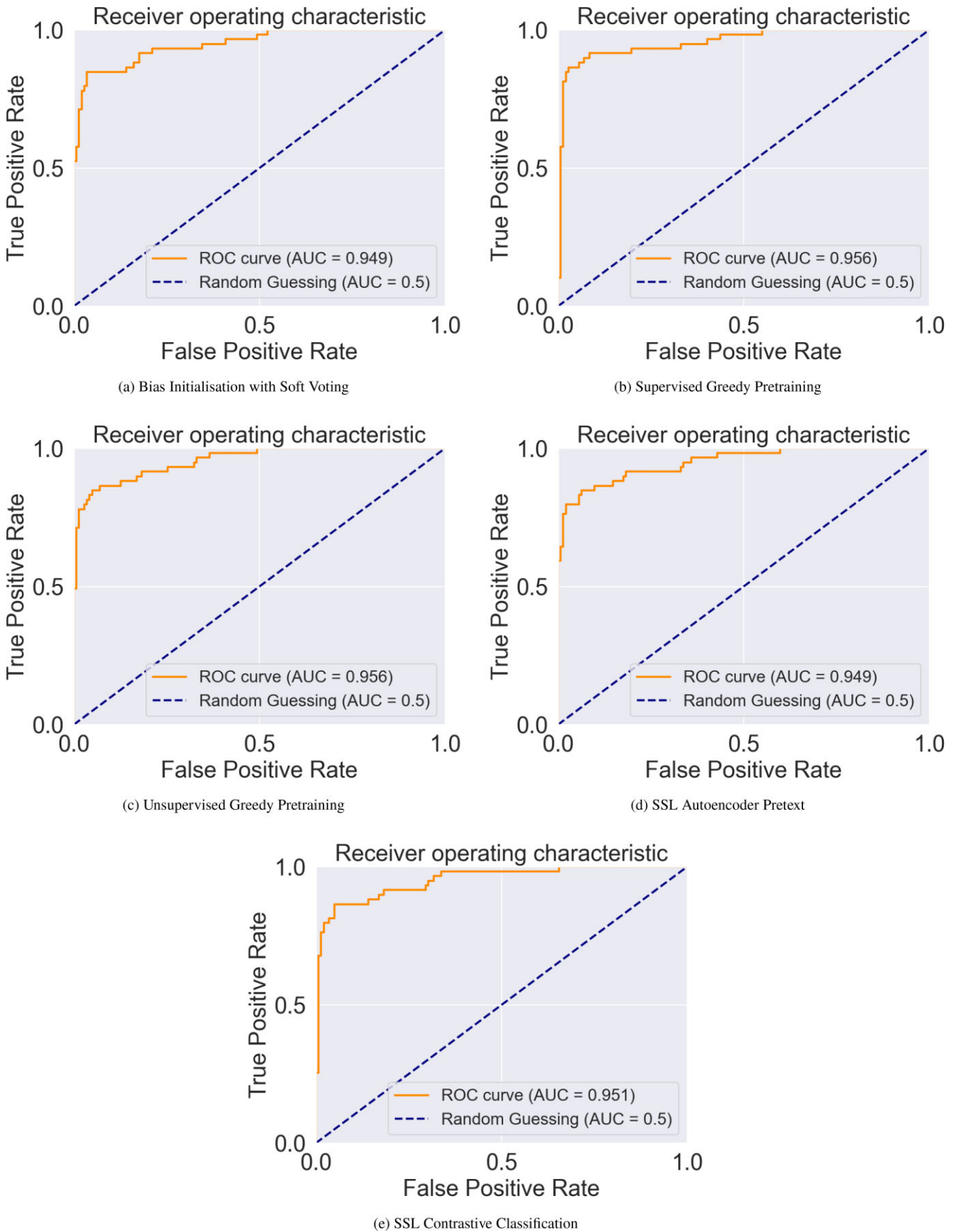


Figure 3. ROC curves along with corresponding AUC values. The ROC curve is plotted between the False Positive Rate and True Positive Rate.

Table 2. Performance summary: a comparison table between macro and weighted averages for all the models employed in this study.

Model	Macro average			Weighted average		
	Precision	Recall	<i>F1</i> score	Precision	Recall	<i>F1</i> score
Bias initialization with soft voting	0.924	0.906	0.914	0.930	0.930	0.930
Supervised greedy pretraining	0.891	0.915	0.901	0.920	0.915	0.917
Unsupervised greedy pretraining	0.857	0.879	0.866	0.891	0.886	0.887
SSL autoencoder pretext	0.863	0.883	0.872	0.895	0.891	0.892
SSL contrastive classification	0.857	0.879	0.866	0.891	0.886	0.887

to have an appropriate metric for comparison, with the help of confusion matrix, we calculate various parameters such as Precision, Recall, *F1* score, and Accuracy as shown in Table 1. As we are doing a binary classification, we also calculate the area under the curve (AUC) score corresponding to every model as seen in the last column of Table 1. To calculate AUC score, we first plot the receiver operating characteristic (ROC) curve that is a plot between the ‘True Positive Rate’ and the ‘False Positive Rate’ as seen in Fig. 3. Next, we calculate the area under curve to identify the AUC value. Ideally, the more close the value is to 1, the better the score is. Though from AUC score it seems that greedy pretraining is an optimal technique to proceed with, the relative difference in the score is minor compared to the misclassification error. Additionally, in the context of imbalanced data classification, *F1* score is a much better metric than AUC. To validate our claim, we further calculate the macro and weighted averages of all the metrics as shown in Table 2. The major difference between macro and weighted average is that macro average is the arithmetic mean of individual scores, while weighted average includes the individual sample sizes. Macro average is calculated using the unweighted mean, which treats all classes equally regardless of their support values. This can penalize the model if the performance in minority classes is poor. On the other hand weighted average takes into account the number of true instances in each class to cope with class imbalance. As seen from the table, both the macro and weighted average *F1* scores is better for the bias initialization with soft voting. Thus considering its lowest misclassification rate and highest weighted and macro *F1* scores, the ‘bias initialization with soft voting’ model turns out to be the best one.

Next, we also plot the histogram corresponding to the output value given by the model on the test data. This helps in identifying if the models are overconfident in their predictions by simply comparing its histogram with the misclassification rate. As in our case BL Lacs are denoted by 0 and FSRQs are denoted by 1, the magnitude of the bars close to 0 relates to the algorithm being confident about the target being BL Lac, similar to the case with FSRQs wherein we focus on magnitudes of the bars close to 1. As seen in Fig. 4(a), compared to its counterparts, model does not give a confident prediction for a large number of samples. It is highly probable that most of the FSRQs misclassified as BL Lacs will have a prediction value around 0.5, and the same is expected to be the case for BL Lacs misclassified as FSRQs. On the other hand for all the other algorithms particularly for the self-supervised ones, we see a very high magnitude bar around the value 0, indicating that model is confident about most of the samples being BL Lacs, and the same has been observed from the confusion matrix. However, an important point to note here is that, these algorithms tend to classify a lot of BLLs as FSRQs and hence being overconfident about wrong predictions. Thus, even though the flatter distribution in Fig. 4(a) might seem counterintuitive to well separated categories, it reflects the fact that this model is not disproportionately biased to either class. The comparatively higher

peaks around 0 for models (Fig. 4c–e) indicate that they are biased towards BLLs, since they identify a considerably lesser number of BLLs correctly than Fig. 4(a).

In addition to its optimal performance, one of the major advantages of our ‘Bias Initialization with Soft Voting Neural Network’ is the model’s parameter count, as discussed in the previous section. Recently, numerous studies have explored similar approaches (Agarwal 2023; Cooper et al. 2023; Sahakyan, Vardanyan & Khachatryan 2023), employing a large number of features and are thus, limited to a small number of samples, due to a high percentage of missing values in the catalogue. On the other hand, if the algorithms estimate these missing parameters e.g. redshift before classification, then their results will be heavily biased based on the imputation technique used. In contrast, Agarwal’s optimal model shares the same number of features as ours. However, the considered model in the work requires approximately seven times more FLOPs compared to the presented model. This computational efficiency enhances the user experience when using our tool without compromising on results. One can find a link to the web app in the DATA AVAILABILITY section.

We make use of each model to make predictions for BCUs.⁶ The histograms can be seen in Fig. 5. Out of 1115 BCU samples, the optimal algorithm (‘bias initialization with soft voting’) of ours predicts that there are 820 BL Lacs and 295 FSRQs, the list of which is made available in our GitHub repository. This number can be considered approximate due to the uncertainty and errors associated with every ML model. One of the other supporting evidence for these numbers is the redshift associated with the BCU samples. In our previous paper, we observed that a significant number of BCUs in the catalogue have a lower redshift (refer to figs 4 and 5 in Gharat, Borthakur & Bhatta 2023), indicating that there would be more BL Lacs compared to FSRQs, which perfectly aligns with the claim made in this study. Table 3 presents the predictive statistics for the unclassified BCUs according to each model. Notably, there is a separate consensus between the greedy pretraining approaches and between the self-supervised learning techniques, with each of the five models concurring that no less than 60 per cent of these samples are likely BLLs.

We observe a significant dominance of BL Lacs in the resulting sample, making up approximately 73 per cent, which is consistent with the observation in the 4LAC catalogue, where the number of BL Lacs is nearly double that of FSRQs. This dominance is further supported by the challenges posed by a large number of BL Lac objects displaying weak or no emission lines, making the detection of their optical counterpart information difficult and classification a complex task. Consequently, it is reasonable to infer that a significant fraction of the BCUs in the *Fermi*/LAT catalogue are highly likely to be BL Lac objects. Also, it is important to note that considering the

⁶https://github.com/abhimanyu911/bcu-classification/tree/main/bcu_predictions

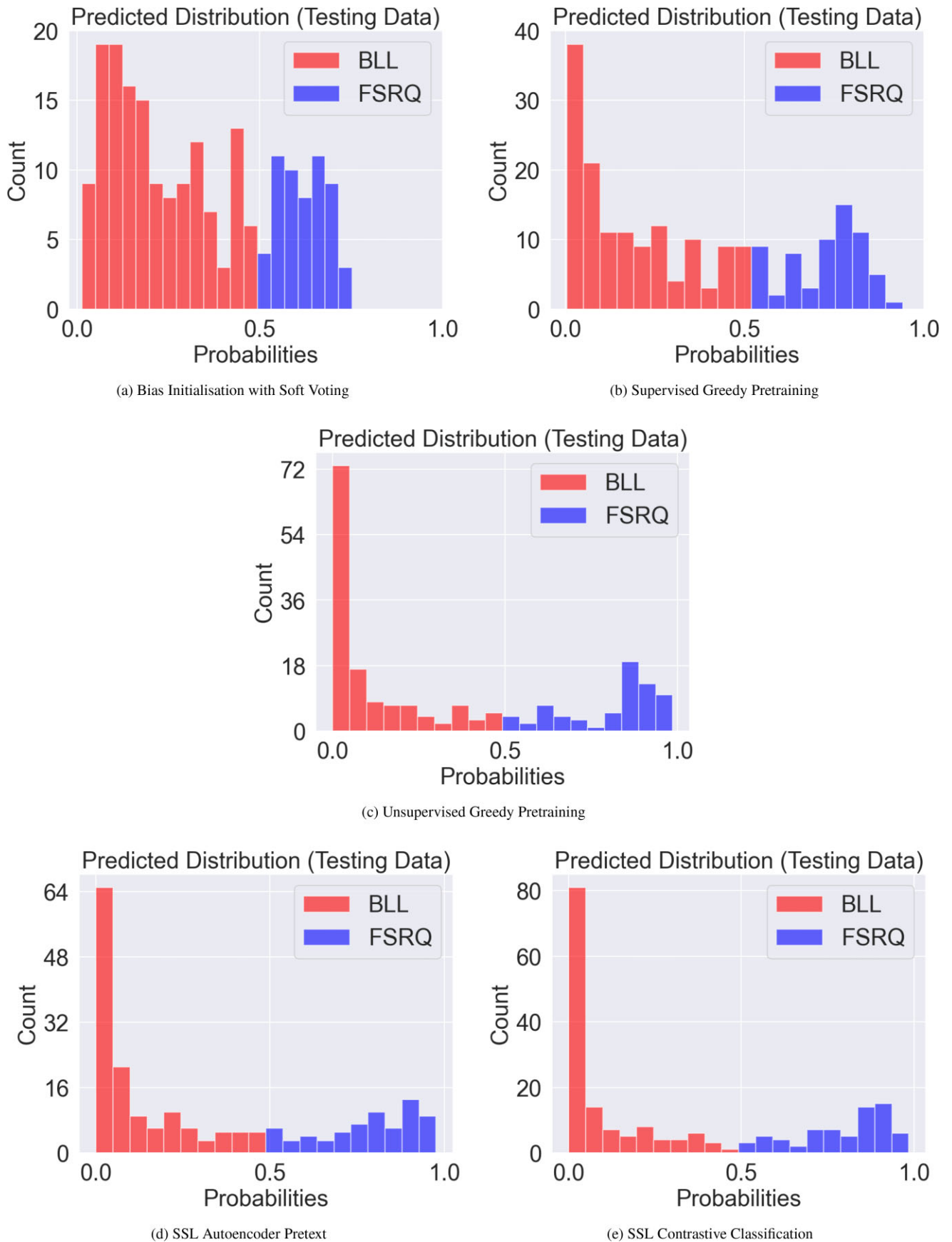


Figure 4. Distribution of predicted output on test data. Here, ‘0’ denotes the BL Lac and ‘1’ denotes the FSRQ. The closer the value is to these extremes, the more confident the model is in its prediction.

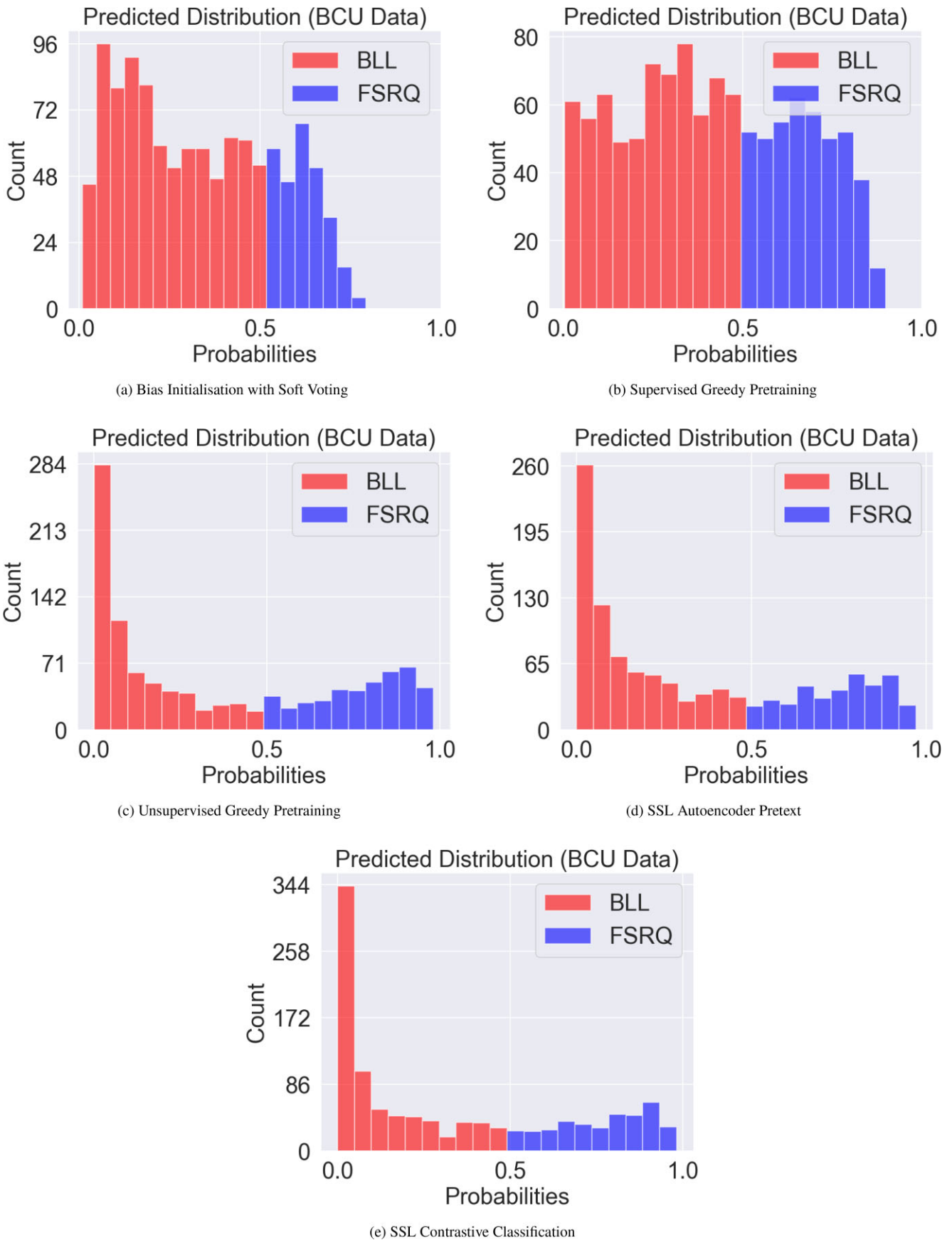
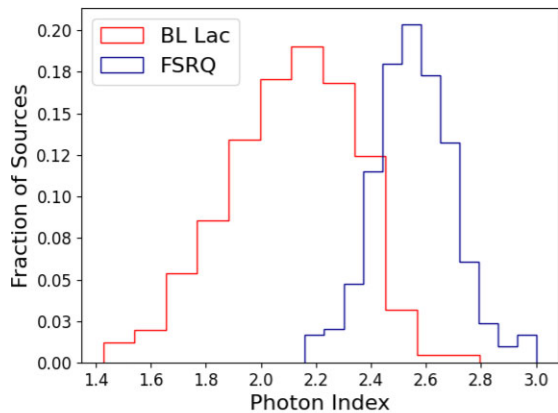


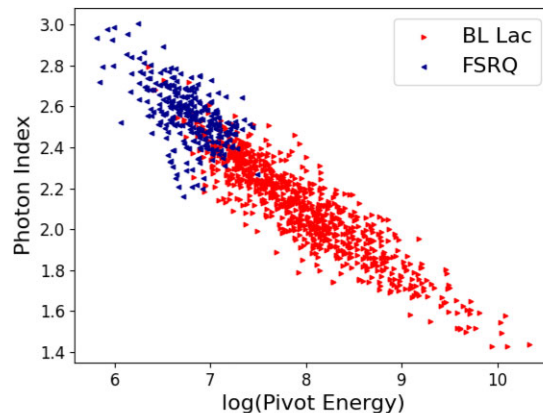
Figure 5. Distribution of predicted output on BCU sources. Here, ‘0’ denotes the BL Lac and ‘1’ denotes the FSRQ. Again, the closer the value is to these extremes, the more confident the model is in its prediction.

Table 3. Number of BLLs and FSRQs predicted for the 1115 BCUs by different models.

Model	Number of BLLs predicted	Number of FSRQs predicted
Bias initialization with soft voting	890	295
Supervised greedy pretraining	691	424
Unsupervised greedy pretraining	691	424
SSL autoencoder pretext	754	361
SSL contrastive classification	757	358



(a) Distribution of Photon Indices in the Predicted Sample



(b) Photon Index Distribution across Logarithmic Pivot Energies in the predicted Sample

Figure 6. The dichotomy between BL Lacs and FSRQs is evident in their distribution within the predicted sample.

binary nature of the classification, limited to BL Lacs and FSRQs, and excluding other potential classes such as Seyferts, radio galaxies, and other AGNs – constituting only a small fraction of the sample – we anticipate that less than 3 per cent of non-blazar AGN subclasses could potentially introduce contamination in the BCU sample.

To explain the apparent dichotomy between BL Lacs and FSRQs, the intrinsic difference in the nature of the accretion disc and physical origins of gamma-ray emissions in BL Lacs and FSRQs may contribute to their distinct gamma-ray properties, such as the larger gamma-ray luminosity of FSRQs and the harder gamma-ray spectral characteristics of BL Lacs. For instance, BL Lacs can have a *radiatively inefficient accretion flow* and be more magnetically dominated, resulting in a lower mass accretion rate compared to FSRQs (e.g. Ghisellini, Maraschi & Tavecchio 2009; Ghisellini et al. 2011; Mondal & Mukhopadhyay 2019). Similarly, the absence of circumnuclear gas near the central engine, as indicated by the featureless optical spectra of BL Lacs, suggests that the origin of gamma-ray emission in the sources can predominantly be ascribed to SSC emission (e.g. Mastichiadis & Kirk 1997). On the other hand, studies of the broad-band SED of FSRQs suggest that gamma-ray emission in FSRQs is mainly contributed by External-Compton (EC) processes (see e.g. Hayashida et al. 2015, and reference therein).

To ensure the model’s consistency in accordance with the characteristic features of the two classes, we plotted the distribution of the photon index. We observed that the BL Lacs in the predicted sample exhibit a relatively harder photon index compared to FSRQs, as illustrated in Fig. 6(a) (see also figs 1 and 3 in Ajello et al. 2022). Additionally, we also plotted a graph between photon index and log of pivot energy and observed results align with the expected trend, positioning BL Lacs in the upper left of the anticorrelation trend, as seen in Fig. 6(b) (see also fig. 2 in Kang et al. 2019). The results from the study can be further verified and refined through dedicated multiwavelength observations, involving telescopes ranging from

radio to TeV. Moreover, as BL Lacs represent the most extreme class of AGNs, this sample can serve as the target sample for future ground-based TeV and PeV telescopes.

4 CONCLUSION

The recent catalogue from the *Fermi*/LAT gamma-ray telescope contains a large number of AGN sources that require decisive classification. This is because classification using gamma-ray data alone is not always possible. In this study, we employ multiple algorithms to classify blazars of unknown class into BL Lacs and FSRQs. The proposed models exhibit simplicity in their nature, with their distinctiveness stemming from the fine-tuning of the method through appropriate initialization and the incorporation of soft voting. Additionally, we delve into a couple of self-supervised algorithms in their vanilla form to assess their capabilities. ‘Bias initialization with soft voting’ emerges as the best-performing model in our case. The algorithm’s ability to make a larger number of predictions can be attributed to the minimal number of features it utilizes, which stands as a key factor. Furthermore, our study places emphasis on maintaining a minimal number of parameters while still delivering strong performance, which proves to be a pivotal feature. This enables us to deploy the model in various scenarios.

ACKNOWLEDGEMENT

We thank the anonymous referee for their careful and thorough review of this paper, which helped improve the quality of the work.

DATA AVAILABILITY

The data utilized in this paper can be accessed by the public through the Fermi Science Support Center of NASA’s Goddard

Space Flight Center. Following the reproducibility and open source standards followed by the ML community, we make all our codes public that can be accessed through our Github repository (<https://github.com/abhimanyu911/bcu-classification>). Further, to have a comfortable experience with our method, the best-performing model is deployed on AWS (<http://13.239.10.157:8501/>) and Streamlit (<https://bcu-classification-ml.streamlit.app/>). Note that the AWS app may get deactivated after the expiry of credits, in such cases we recommend a user to make use of Streamlit. In case of any issues, a docker image can be provided on reasonable request to sarveshgharat19@gmail.com.

REFERENCES

- Abdollahi S. et al., 2020, *ApJS*, 247, 33
 Agarwal A., 2023, *ApJ*, 946, 109
 Ajello M. et al., 2022, *ApJS*, 263, 24
 Atwood W. B. et al., 2009, *ApJ*, 697, 1071
 Bae K., Ryu H., Shin H., 2019, preprint ([arXiv:1911.00289](https://arxiv.org/abs/1911.00289))
 Bhatta G., 2021, *ApJ*, 923, 7
 Bhatta G., Dhital N., 2020, *ApJ*, 891, 120
 Bhatta G., Mohorian M., Bilinsky I., 2018, *A&A*, 619, A93
 Błażejowski M., Sikora M., Moderski R., Madejski G. M., 2000, *ApJ*, 545, 107
 Bloom S. D., Marscher A. P., 1996, *ApJ*, 461, 657
 Böttcher M., 2019, *Galaxies*, 7, 20
 Butter A., Finke T., Keil F., Krämer M., Manconi S., 2022, *J. Cosmol. Astropart. Phys.*, 2022, 023
 Chen X. et al., 2023, *Int. J. Comput. Vis.*, 1
 Cooper N., Dainotti M. G., Narendra A., Lioudakis I., Bogdan M., 2023, *MNRAS*, 525, 1731
 Dermer C. D., Schlickeiser R., 1993, *ApJ*, 416, 458
 Erhan D., Courville A., Bengio Y., Vincent P., 2010, in Yee Whye T., Mike T., Whye Y., eds, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. PMLR, p. 201
 Fossati G., Maraschi L., Celotti A., Comastri A., Ghisellini G., 1998, *MNRAS*, 299, 433
 Gharat S., Borthakur A., Bhatta G., 2024, *MNRAS*, 527, 6198
 Ghisellini G., Maraschi L., Tavecchio F., 2009, *MNRAS*, 396, L105
 Ghisellini G., Tavecchio F., Foschini L., Ghirlanda G., 2011, *MNRAS*, 414, 2674
 Ghisellini G., Righi C., Costamante L., Tavecchio F., 2017, *MNRAS*, 469, 255
 Hayashida M. et al., 2015, *ApJ*, 807, 79
 Henaff O., 2020, in *International Conference on Machine Learning*. PMLR, p. 4182
 Jorstad S. G. et al., 2017, *ApJ*, 846, 98
 Kang S.-J., Li E., Ou W., Zhu K., Fan J.-H., Wu Q., Yin Y., 2019, *ApJ*, 887, 134
 Kovačević M., Chiaro G., Cutini S., Tosti G., 2020, *MNRAS*, 493, 1926
 Liu H., Jiang X., Li X., Guo A., Hu Y., Jiang D., Ren B., 2023, in *Proceedings of the AAAI Conference on Artificial Intelligence*. p. 1649
 Maraschi L., Ghisellini G., Celotti A., 1992, *ApJ*, 397, L5
 Mastichiadis A., Kirk J. G., 1997, *A&A*, 320, 19
 Mehta S., Paunwala C., Vaidya B., 2019, in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*. IEEE, p. 1293
 Mondal T., Mukhopadhyay B., 2019, *MNRAS*, 486, 3465
 Sahakyan N., Vardanyan V., Khachatryan M., 2023, *MNRAS*, 519, 3000
 Sikora M., Begelman M. C., Rees M. J., 1994, *ApJ*, 421, 153
 Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R., 2014, *J. Mach. Learn. Res.*, 15, 1929
 Urry C. M., Padovani P., 1995, *PASP*, 107, 803
 van den Oord A., Li Y., Vinyals O., 2018, preprint ([arXiv:1807.03748](https://arxiv.org/abs/1807.03748))

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.