Machine Learning for Real Time Classification of Transient Events: a Recurrent Neural Network Auto-Encoder and Gradient Boosting Classifier

DANIELLE YANG
INDIANA UNIVERSITY BLOOMINGTON LUDDY SCHOOL OF INFORMATICS

The rapid advancement of astronomical survey technologies, such as the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST), is expected to generate millions of transient events annually, posing significant challenges in processing large volumes of unlabeled data. To address this, a deep learning model was developed, combining a Recurrent Neural Network Variational Autoencoder (RNN-VAE) for dimensionality reduction with a Gradient Boosting Classifier for real-time classification of transient events. This model efficiently classifies galactic and extragalactic transients without the need for labeled data. Using the PLAsTiCC dataset, the model achieved an AUC-ROC score of 0.94 and F-1 score of 0.89, demonstrating strong performance in distinguishing between various transient classes, including rare events. This approach offers a scalable solution for real-time astronomical surveys, enhancing both classification accuracy and resource allocation in future data-rich environments.

DOI: https://doi.org/10.60690/5rzzwg14

Introduction

The rapid advancement of astronomical survey technologies is set to provide astronomers with unprecedented volumes of data. The Vera C. Rubin Observatory's decade-long Legacy Survey of Space and Time (LSST), in particular, is expected to detect millions of transient events annually (Ivezić et al., 2019). This data influx offers the potential for the discovery of rare and previously unknown transient events.

However, managing this overwhelming volume of data presents significant challenges. Identifying which transients merit further investigation becomes increasingly difficult with the scale of incoming data. Spectroscopic follow-up, a critical method for understanding transient events, is both resource-intensive and requires a meticulous approval process, further complicating the selection of optimal targets for observation. Additionally, transient events must be captured near their peak brightness to justify the allocation of these limited resources.

To address these challenges, efficient methods for real-time labeling and prediction of transient events are urgently required. Early efforts in this field focused on photometric classification to build pure samples of transients (Boone, 2019). More recent approaches have shifted towards anomaly detection, with several methods demonstrating success when using complete light curve data from extensive surveys (Pasquet et al., 2018; Muthukrishna et al., 2019). However, these models struggle when faced with incomplete data, a common challenge in real-time analysis. To overcome this limitation, recent studies have explored the use of autoencoders to manage unlabeled data (Morawski et al., 2021).

This paper proposes a model aimed at real-time classification of galactic and extragalactic transient events, building on recent advancements in deep learning and autoencoder-based architectures. Section two describes the dataset and feature extraction process, while section three, four, and five outline the model and training process of the RNN-VAE and Gradient Boosting Classifier (GBC). The results are discussed in section six, where the model is heavily evaluated and compared to baseline models, and its limitations and future work are further stated. Finally, the paper is concluded in section 7.

Data

The model was trained and evaluated using the Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC) dataset, which was initially created for a public competition hosted on Kaggle. Although the competition has since concluded, the dataset and the methods it inspired remain highly relevant due to the scale and innovation it introduced to the data science community (Hložek et al., 2020). The PLAsTiCC dataset contains simulated data based on approximately three years of observation from the Rubin Observatory (Kessler et al., 2019). It includes light curves and metadata such as redshift, flux, magnitude, corresponding errors, and additional physical parameters across six observational passbands (ugrizy). For this model, light curves and metadata were integrated

and processed through the RNN-VAE for feature extraction prior to classification.

For this study, 100,000 data points spanning sixteen transient event classes were used. These classes were selected to provide a balanced representation of both common and rare astronomical events. Frequently occurring events, such as supernovae, were well-represented, allowing the model to effectively learn their distinguishing characteristics. Conversely, rarer events like Tidal Disruption Events (TDEs) and Calcium-rich transients (CaRTs) were included to assess the model's capability in classifying minority classes, which typically present more challenges due to their limited instances.

To ensure balance, event classes with fewer than three instances were excluded from the analysis, as they would not provide sufficient data for reliable classification. This selection process enabled the model to be exposed to a diverse, yet representative, range of transient phenomena, ensuring it could generalize well across different event types. Balancing the dataset was crucial in mitigating bias toward overrepresented classes, while still allowing the model to effectively handle underrepresented, yet scientifically significant, events. Further details of the data selection process, including class distributions and the methods used to preserve their frequencies, are discussed in the pre-processing section.

AGN (Active Galactic Nuclei)
Active Galactic Nuclei (AGN) are highly energetic regions located at the centers of galaxies, characterized by the emission of intense radiation and variability in brightness over different timescales. Despite being relatively common, AGNs make up approximately 6.3% of the dataset, aligning with their prevalence in the universe.

CaRT (Calcium-Rich Transients)
Calcium-rich transients are a rare type of supernova known for their rapid evolution, relatively low peak luminosity, and strong calcium emission lines. These events account for roughly 0.3% of the dataset, accurately reflecting their rarity in astronomical observations.

EB (Eclipsing Binaries)
Eclipsing Binaries are systems of two stars whose orbits are oriented such that they periodically eclipse each other from the perspective of Earth. This leads to fluctuations in brightness. Given their common occurrence, EBs are well-represented in the dataset.

ILOT (Intermediate Luminosity Optical Transients)
Intermediate Luminosity Optical Transients (ILOTs) are astronomical events that explode with brightness levels between novae and supernovae. These rare occurrences account for less than 0.1% of the dataset.

M-Dwarf (Low-Mass Stars)
M-dwarfs, also known as low-mass stars, are stars smaller and significantly less luminous than the Sun. These stars represent a substantial portion of the dataset due to their frequency in the universe.

Mira (Mira-Type Variable Stars)
Mira-type variable stars are cool red giants known for their periodic changes in brightness. They are massive and very luminous compared to the Sun, although they make up less than 0.1% of the dataset, reflecting their status as a minority event.

PISN (Pair Instability Supernovae)
Pair Instability Supernovae (PISN) are rare types of supernovae that result from collisions between gamma rays and atomic nuclei, producing free electrons and positrons. These events also make up less than 0.1% of the dataset.

RRL (RR Lyrae Variable Stars)
RR Lyrae variable stars are periodic variable stars commonly found in globular clusters. They make up a significant portion of the dataset due to their well-understood light curve characteristics.

SLSN-I (Type I Superluminous Supernovae)
Type I Superluminous Supernovae (SLSN-I) are hydrogen-free stellar explosions that are roughly ten times more luminous than typical supernovae. These events constitute approximately 2% of the dataset.

SN-Ia (Type Ia Supernovae)
Type Ia Supernovae (SN-Ia) are binary systems containing a white dwarf star. Known for their consistent peak luminosity and lack of hydrogen in their spectra, they represent over half of the dataset, making them the most frequent class of transient events.

SN-II (Type II Supernovae)
Type II Supernovae (SN-II) are massive stellar explosions characterized by the retention of hydrogen. These events, which exhibit long plateaus in their light curves, account for about one-fourth of the dataset, making them a major class.

SNIa-91bg (Type Ia-91bg Supernovae)
These are dimmer, faster, and redder variants of Type Ia Supernovae. Due to their shorter observable window, they make up about 1% of the dataset.

SNIax (Type Iax Supernovae)
Type Iax Supernovae are dimmer and faster than traditional SN-Ia events, although they are not necessarily redder. They constitute roughly 2% of the dataset.

SNIbc (Type Ibc Supernovae)
Type Ibc Supernovae result from the explosions of stars that have shed their helium and hydrogen envelopes. These events account for approximately 6% of the sample.

TDE (Tidal Disruption Events)
Tidal Disruption Events (TDEs) occur when a star is torn apart by the immense gravitational forces of a supermassive black hole. These rare events make up less than 1% of the dataset.

muLens-Single (Gravitational Microlensing Events)
Gravitational microlensing events result in a temporary increase in the brightness of background stars due to the gravitational influence of a massive foreground object. These events are uncommon, comprising less than 2% of the dataset.

Data Pre-Processing:

Before the light curves were input into the model, a thorough pre-processing routine was conducted to ensure the quality and reliability of the training data. Initially, any data points exceeding seven standard deviations from the mean were removed to

mitigate the impact of extreme outliers. Following this, sigma clipping was applied for five iterative cycles to account for potential shifts in the mean caused by residual outliers.

Several additional steps were implemented to address the astrophysical characteristics of the dataset. For nearby galactic events, particularly those with redshift values around 0.001, the distance modulus was adjusted to reflect a standard distance of approximately 10 megaparsecs (Heymann, 2012). This adjustment ensured consistency in distance measurements across different events, facilitating more accurate comparisons.

The magnitude scale was then recalibrated, with brightness increases represented by smaller apparent magnitudes, in line with standard astronomical observational practices. Additionally, to correct for dust extinction effects in the Milky Way, the Fitzpatrick (1999) extinction model was applied. This model, based on Milky Way parameters, assumes an extinction law of Rv = 3.1, with central wavelengths corresponding to the six LSST filters (ugrizy).

Event classes with fewer than three instances were excluded from the dataset to enhance the model's performance and ensure that statistically significant data were included in the final analysis. This pre-processing phase was essential for preparing the dataset, ensuring that it was both clean and representative of the various transient events being classified.

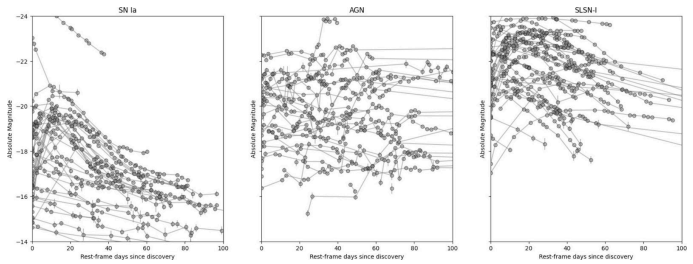

Figure 1: These plots graph the pre-processed lightcurves of SN Ia, AGN, and SLSN-I in terms of absolute magnitude. Note: created by student researcher

RNN-VAE Model Architecture

The model architecture employed in this study integrates a Recurrent Neural Network Variational Autoencoder (RNN-VAE) to reduce data dimensionality while retaining critical features of transient events. This architecture was chosen based on previous findings that demonstrated the effectiveness of feature extraction in supernova classification (V. Villar et al., 2019). The RNN-VAE model combines both RNN and VAE components, providing a dual advantage: capturing temporal dependencies in the data while generating a meaningful latent space representation.

The RNN-VAE comprises two main components: the encoder and the decoder (Figure 2). The encoder processes light curves, generating parameters that define the data's latent space representation.

Meanwhile, the decoder reconstructs the light curves from this compressed representation, reducing the feature set while maintaining vital information.

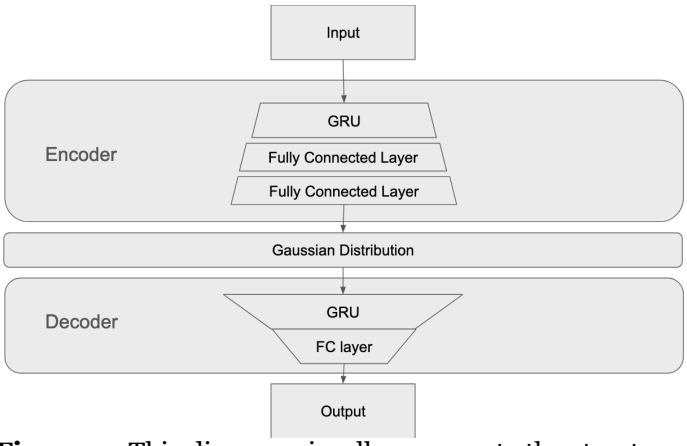

Figure 2: This diagram visually represents the structure of the RNN-VAE. Note: created by student researcher.

Encoder:
The encoder compresses the input light curves into a Gaussian distribution. It operates in three stages:

1. Gated Recurrent Unit (GRU): The input data passes through a GRU, a type of recurrent neural network that leverages sigmoid and hyperbolic tangent (tanh) activation functions to transform the input into latent outputs. These activation functions help regulate the information flow, deciding how much to retain from previous steps.
2. Fully Connected Layers: Two fully connected layers follow, applying linear and non-linear transformations to learn complex relationships between features.
3. Latent Space Representation: The final output of the encoder comprises the mean, standard deviation, and Gaussian distribution parameters that define the latent space.

Decoder:
The decoder reverses the encoding process, reconstructing light curves from the reduced latent space representation. It uses additional GRUs and fully connected layers to map the compressed data back to its original form. The decoder's output sequence retains only the most relevant features, optimized to include the six most significant features necessary for classification .

Feature Selection and Optimization:
The decision to limit the number of features to six was based on performance trials aimed at balancing model complexity and classification accuracy. These trials tested various feature counts, ranging from three to twelve, evaluating both the accuracy of the final classifier and the training/validation loss of the RNN-VAE itself. Results indicated that fewer than six features led to a drop in accuracy due to the loss of essential information, while more than six features resulted in diminishing returns and increased complexity, causing overfitting. Six features were thus selected as the optimal number for maintaining detail while ensuring model efficiency.

Table 1: Number of Features vs. Accuracy

Number of Features	Accuracy
1	0.001
2	0.03
3	0.15
4	0.56
5	0.79
6	0.89
7	0.88
8	0.89
9	0.82
10	0.83
15	0.78
20	0.82
25	0.65

Table 1: This table of values compares the number of features to the overall model's accuracy, demonstrating 6 features as the most accurate and efficient number. Note: created by student researcher.

How GRU Works in RNN-VAE:

The Gated Recurrent Unit (GRU) is a critical component of the RNN-VAE architecture, designed to efficiently process sequential data, such as the light curves of transient events. It is a type of recurrent neural network that optimizes the handling of temporal dependencies while minimizing computational complexity. Unlike traditional Recurrent Neural Networks (RNNs), GRUs utilize gating mechanisms to control the flow of information, making them particularly suited for time-series data with irregular time steps or missing values—common in astronomical datasets.

The GRU's structure (Figure 3) is simpler than that of the Long Short-Term Memory (LSTM) units but retains similar effectiveness in sequence learning. It comprises two primary gates—an update gate and a reset gate—that regulate the flow of information through the network:

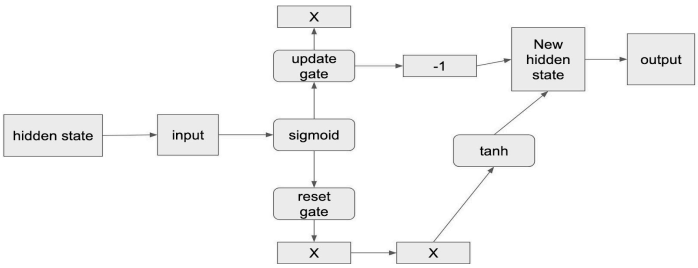

Figure 3: This diagram visually represents the structure of a GRU. Note: created by student researcher.

Update Gate:

- The update gate determines how much of the previous information needs to be carried forward to the next time step.
- It combines the roles of the LSTM's forget and input gates, effectively deciding which information to retain and which to discard.
- Mathematically, the update gate (ztz_tzt) is calculated as:
  $z_t = \sigma(W_z \bullet [h_{t-1}, x_t])$
  where W_z is the weight matrix, h_{t-1} is the hidden state from the previous time step, and x_t is the current input.

Reset Gate:

- The reset gate decides how much of the past information should be forgotten at the current time step.
- This allows the GRU to focus more on recent information when making predictions, which is crucial for reconstructing light curves in the decoder.
- The reset gate (r_t)is computed as:
  $r_t = \sigma(W_r \bullet [h_{t-1}, x_t])$

Current Memory Content:

After determining the update and reset gates, the GRU computes the candidate hidden state ($\widehat{h_t}$), which incorporates both the current input and the previous hidden state,

t modulated by the reset gate: $\widehat{h_t} = tanh(W \bullet [r_t * h_{t-1}, x_t])$

Here, the reset gate (r_t) controls how much of the past information influences the current candidate hidden state.

Hidden State Update:

The final hidden state h_t is a linear interpolation between the previous hidden state and the candidate hidden state, determined by the update gate:

$$h_t = (1 - z) * h_{t-1} * \widehat{h_t}$$

Role of GRU in RNN-VAE for Transient Classification:

The GRU's gating mechanisms allow the model to selectively retain or discard information at each time step, making it highly effective in modeling sequential dependencies in light curves. In the context of transient event classification, GRUs help capture key temporal patterns in the evolving brightness of different astronomical phenomena. This capability is essential for accurately reconstructing light curves and extracting meaningful features for subsequent classification.

The GRU's simplicity and efficiency, compared to more complex units like LSTMs, also make it faster to train while maintaining strong performance. This balance between computational efficiency and temporal accuracy is critical for handling the large, real-time datasets generated by astronomical surveys like the LSST.

Training and Optimization

The training of the Recurrent Neural Network Variational Autoencoder (RNN-VAE) was carried out in distinct phases: initialization, the training loop, optimization, and validation.

Initialization Phase:
Initially, the model's parameters were randomly assigned to ensure reproducibility across trials. Random initialization is crucial for establishing a baseline and enabling consistent performance comparisons (Kingma & Ba, 2015).

Training Loop:
The model was trained over multiple epochs, with each epoch comprising batches of data, allowing efficient processing of the dataset (Paszke et al., 2019). The training continued until either the entire dataset was processed or improvements in performance stagnated. Each epoch involved a forward pass through the encoder and decoder, followed by a backward pass to compute gradients and update model parameters.

During the forward pass:
- Encoder: The encoder compresses input light curves into latent space representations (Rezende et al., 2014).
- Decoder: The decoder reconstructs the light curves from this latent space.

After the forward pass, the backward pass computed gradients and loss values, which were used to refine model parameters and improve accuracy.

Optimization Strategy:
The optimization process employed the Adaptive Moment Estimation (Adam) optimizer, selected for its ability to adjust learning rates dynamically based on the first and second moments of the gradients (Kingma & Ba, 2015). Adam is particularly effective for managing sparse gradients in deep learning models, making it suitable for sequential data processing in this context.
The optimization process involved:
1. Gradient Calculation: Gradients of the loss were computed with respect to model parameters.
2. Parameter Updates: Parameters were iteratively updated to minimize the loss, facilitating convergence.

Composite Loss Function:
To maximize model performance, a composite loss function was utilized:
1. Mean Squared Error (MSE): Measures the reconstruction error between the original and reconstructed light curves, ensuring critical features are retained (Bengio et al., 2013).

$$MSE = \frac{1}{n}\sum (y - y_1)^2$$

2. Kullback-Leibler (KL) Divergence: Measures the divergence between the latent distribution and a Gaussian prior, guiding the latent space to conform to a normal distribution (Kingma & Welling, 2013).

$$KL(P||Q) = \sum P(X) \times log\frac{P(X)}{Q(X)}$$

The combination of MSE and KL Divergence aimed to maximize the Evidence Lower Bound (ELBO), a standard objective in VAEs that improves both reconstruction quality and the latent space distribution (Rezende et al., 2014).

Hyperparameter Tuning for GRU:
Several hyperparameters were adjusted throughout training to optimize model performance:
- **Number of Layers:** The GRU layers were set to 2, balancing complexity and computational efficiency (Cho et al., 2014).
- **Hidden Units:** 64 hidden units per GRU layer were chosen based on validation accuracy, ensuring sufficient capacity to capture light curve features.
- **Learning Rate:** A starting learning rate of 0.001 was used, with gradual decay to prevent overshooting minima during optimization.
- **Batch Size:** A batch size of 128 was found to be optimal, offering a good balance between convergence speed and stability (Paszke et al., 2019).

Validation and Early Stopping:
Validation was performed after each epoch, with validation loss serving as the primary performance metric. Early stopping was implemented to prevent overfitting, terminating training when the validation loss began to increase while the training loss remained stable (Prechelt, 1998). This comprehensive training process enabled the model to effectively extract meaningful features from light curves while maintaining robustness in classification tasks (Villar et al., 2019).

Gradient Boosting Classifier

A Gradient Boosting Classifier is an ensemble machine learning model that iteratively trains multiple sequential models to mitigate the errors of each preceding layer. The classifier utilizes a loss function known as deviance loss, which is derived from the negative log-likelihood of the multinomial distribution. Deviance loss imposes significant penalties for misclassifications, making it particularly effective for handling heavily imbalanced datasets, as illustrated in the following equation:

$$Deviance\ Loss = 2\sum[y_1(log)\frac{y_1}{\hat{\mu}} - (y_1 - \hat{\mu})]$$

In addition to employing deviance loss, the Gradient Boosting Classifier incorporates gradient calculations to minimize the loss concerning the outputs of the preceding learner, such as the RNN-VAE utilized in this study. This methodology is akin to the approaches taken by previous researchers who employed an Extreme Boosting Classifier to classify supernovae
based on extracted features (M¨oller et al. 2016).

The base classifier within this framework is a decision tree, often referred to as a "weak learner." However, when combined in a Gradient Boosting Classifier, these decision trees form a much stronger predictive model. Gradient Boosting Classifiers are known for their high predictive capabilities and superior performance in handling heterogeneous data. Given the complex nature of the dataset in this study, the Gradient Boosting Classifier demonstrated improved performance compared to alternative models, such as Random Forest and Support Vector Machines.

The model underwent an iterative fine-tuning process over 105 cycles, during which key hyperparameters were carefully adjusted to optimize performance. These parameters included the number of estimators, maximum tree depth, minimum sample split, and purity decrease threshold. By methodically varying these

settings, the process effectively mimicked a systematic grid search, a common technique used in machine learning to identify the best parameter combinations for enhancing model accuracy.

- **Number of Estimators:** This parameter determines how many individual decision trees are used in the Gradient Boosting Classifier. Increasing the number of estimators typically improves performance but also raises the risk of overfitting. In my fine-tuning, I explored a range of values to find an optimal balance that maximized classification accuracy without compromising the model's generalizability.
- **Maximum Tree Depth:** The maximum depth of each decision tree controls the complexity of the model. A deeper tree can capture more intricate patterns in the data, but it may also lead to overfitting. Through the tuning process, I tested various depths to ensure that the model could adequately learn the underlying structures of the light curves while maintaining robustness against noise.
- **Minimum Sample Split:** This parameter specifies the minimum number of samples required to split an internal node in the decision tree. By adjusting this value, I could control how sensitive the model was to fluctuations in the data. Lower values allowed for more splits and potentially more detailed decision boundaries, while higher values enforced more generalization.
- **Purity Decrease Threshold:** This threshold determines the minimum reduction in impurity required to create a split in the decision tree. Fine-tuning this parameter enabled me to balance between creating too many splits (which could lead to overfitting) and too few (which could result in underfitting).

Throughout the 105 cycles, each adjustment was monitored for its impact on classification performance across a variety of transient event classes. The iterative process allowed me to systematically evaluate how changes in hyperparameters influenced the model's ability to accurately classify different types of transient events, from supernovae to rare events
like Tidal Disruption Events (TDEs).

By carefully adjusting these parameters and observing the corresponding performance metrics, I achieved significant improvements in classification accuracy. The refined model demonstrated enhanced predictive capabilities, effectively capturing the complexities of the light curves associated with each transient event class. This optimization process not only improved the model's performance but also provided valuable insights into the relationships between hyper parameters and model outcomes, informing future research and development efforts.

Results and Discussion

The RNN-VAE model, paired with the Gradient Boosting Classifier (GBC), was trained and tested on the Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC) dataset (Kessler et al., 2019). The model achieved an overall accuracy of 89%, an AUC-ROC score of 0.94, and an average precision, recall, and F-1 score of 0.89, demonstrating strong discriminative performance across a wide range of transient classes.

Class-Specific Performance:

To provide a detailed evaluation of the model's effectiveness, precision, recall, and F1-score were recalculated for each transient class (Figure 4). Given the imbalanced dataset, these metrics offer more meaningful insights than accuracy alone. The following analysis covers the updated performance metrics for each class:

Active Galactic Nuclei (AGN):

The model achieved a precision of 0.88, recall of 0.89, and an F1-score of 0.88 for AGN. Misclassifications were minimal, indicating that the model effectively distinguished AGNs from other classes. The use of advanced feature extraction could further enhance performance.

Calcium-Rich Transients (CaRTs):

The model performed well for CaRTs, achieving a precision of 0.93, recall of 0.94, and an F1-score of 0.93. This strong performance suggests that the model successfully captured the distinct characteristics of CaRTs, despite their limited representation in the dataset.

Eclipsing Binaries (EBs):

The model accurately classified EBs, achieving a precision of 0.96, recall of 0.94, and an F1-score of 0.95. The model's effectiveness in capturing the periodic patterns typical of EBs indicates its robustness in handling cyclic events.

Intermediate-Luminosity Optical Transients (ILOTs):

The model demonstrated high performance for ILOTs, achieving a precision of 0.96, recall of 0.95, and an F1-score of 0.95. Misclassifications were minimal, suggesting that the model captured the distinct characteristics of ILOTs effectively.

M-Dwarf Flares:

The model exhibited strong performance in classifying M-Dwarf Flares, with precision, recall, and F1-score all at 0.96. The model's high performance is attributed to the short-lived flaring behavior of M-Dwarfs, which is easily identifiable in photometric data.

Mira Variables:

The model showed strong performance for Mira Variables, achieving precision, recall, and F1-score of 0.95 each. This performance highlights the model's ability to effectively capture the distinct characteristics of Mira Variables.

Pair-Instability Supernovae (PISN):

The model exhibited high performance for PISN, achieving precision of 0.98, recall of 0.99, and an F1-score of 0.98. The high scores indicate that the model successfully differentiated PISN from other supernova classes.

RRLyrae Variables (RRL):

The model achieved strong performance for RRLyrae Variables, with precision, recall, and F1-score all at 0.95. The model's effectiveness in identifying the periodic variations typical of RRLyrae demonstrates its robustness in handling periodic events.

Type I Superluminous Supernovae (SLSN-I):

The model demonstrated strong performance in classifying SLSN-I, achieving precision of 0.92, recall of 0.94, and an

F1-score of 0.93. The model effectively identified the high-luminosity patterns characteristic of SLSN-I, with occasional misclassifications due to similarities with other supernova classes.

Type Ia Supernovae (SN Ia):
The model achieved a precision of 0.68, recall of 0.61, and an F1-score of 0.64 for SN Ia. Misclassifications were frequent with SN Iax and SN II, primarily due to overlapping early-phase light curve characteristics. Incorporating additional data or feature extraction techniques could improve performance.

Type II Supernovae (SN II):
The model showed moderate performance for SN II, achieving a precision of 0.64, recall of 0.61, and an F1-score of 0.62. Misclassifications primarily occurred with SN Ia, indicating similar early brightness characteristics. Further refinement in feature extraction could improve classification accuracy.

Type Ia-91bg Supernovae (SN Ia-91bg):
The model performed well for SN Ia-91bg, achieving a precision of 0.90, recall of 0.94, and an F1-score of 0.92. The high scores suggest that the model successfully distinguished this subtype from other supernovae.

Type Iax Supernovae (SN Iax):
The model performed well for SN Iax, achieving a precision of 0.85, recall of 0.86, and an F1-score of 0.85. Misclassifications with SN Ia were common, suggesting that further refinement of features could improve differentiation.

SNIbc (Type Ib/c Supernovae):
The model performed well for SNIbc, achieving a precision of 0.81, recall of 0.82, and an F1-score of 0.81. The balanced performance suggests that the model successfully captured the characteristics of SNIbc, although further improvements in recall could enhance accuracy.

Tidal Disruption Events (TDEs):
The model performed well in classifying TDEs, achieving a precision of 0.90, recall of 0.93, and an F1-score of 0.91. The model effectively captured the unique features of TDEs, such as their sharp rise and gradual decay.

muLens-Single:
The model achieved high performance for muLens-Single, with precision of 0.96, recall of 0.99, and an F1-score of 0.97. The model effectively captured the features of gravitational microlensing events, demonstrating its capability to handle rare transient classes.

Overall, the RNN-VAE model, paired with the GBC, demonstrated strong performance across a broad range of transient classes, as indicated by the precision, recall, and F1-scores. While the model excelled in handling imbalanced classes like TDEs and CaRTs, further refinement is needed to improve the classification of overlapping classes, particularly SN Ia, SN Iax, and SN II.

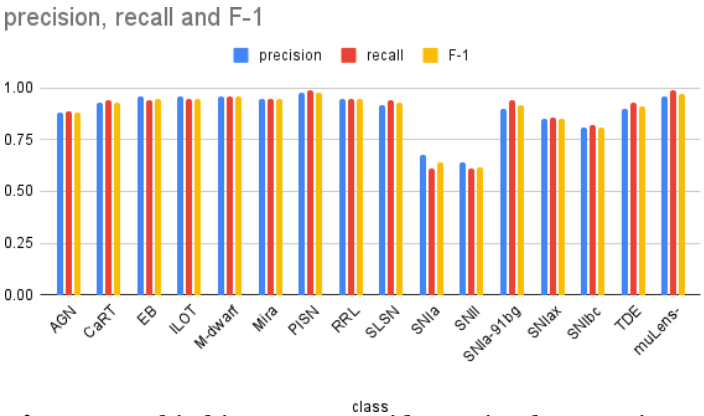

Figure 4: This histogram provides a visual comparison of the classes against each other.

Confusion Matrix Analysis:
The confusion matrix (Figure 5) provided insights into the model's strengths and weaknesses across all 16 classes (Pasquet et al., 2018). While the diagonal elements showed high accuracy for most classes, the highest misclassification rates occurred between SN Ia and SN II, as well as between SN Iax and SN Ia. This suggests that while the model performs well with distinct classes, it struggles with classes that have overlapping photometric features. Enhancing the feature extraction process or incorporating additional data could reduce these errors.

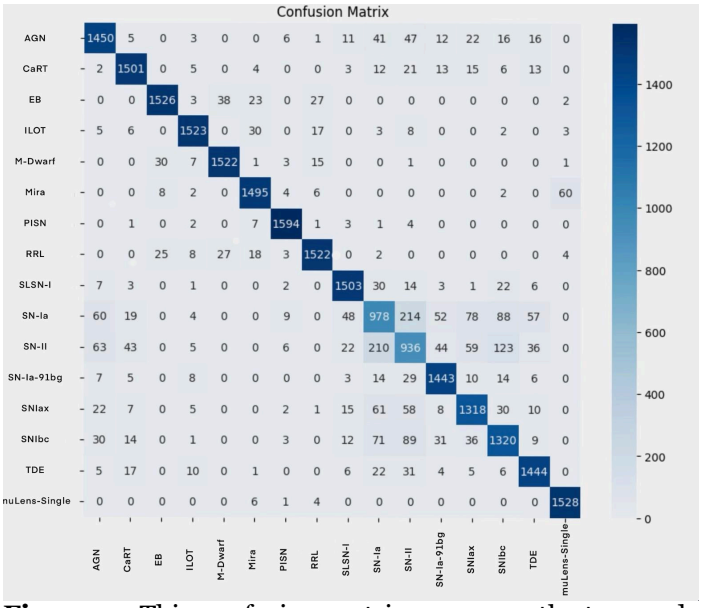

Figure 5: This confusion matrix compares the true and predicted labels of the data set against each other. Ideally, a confusion matrix appears as a dark, diagonal line. Note: created by student researcher.

Handling Class Imbalance:
Class imbalance was a significant challenge during model training, as common classes were more prevalent than rare ones (Boone, 2019). To address this, oversampling and class-weight adjustments were

used. These strategies improved recall for underrepresented classes but introduced a slight increase in false positives, particularly for classes with similar features (Kessler et al., 2019). Future iterations could further refine these methods to improve the balance between precision and recall.

ROC and AUC-ROC Curve:
An average AUC-ROC curve (Figure 6) was generated for all 16 transient classes, providing a comprehensive evaluation of the model's overall performance (Hložek et al., 2020). With an average AUC-ROC score of 0.94, the model demonstrated robust discriminative ability across classes. Averaging the AUC-ROC scores simplified performance comparisons and facilitated a clear assessment of the model's effectiveness in distinguishing transient types, even those with similar light curve features (Villar et al., 2019).

Comparison with Baseline Models:
The RNN-VAE + GBC model was compared against baseline models (Figure 7) used in the original PLAsTiCC competition, such as Random Forests, Support Vector Machines (SVMs), and Convolutional Neural Networks (CNNs) (Pasquet et al., 2018). The RNN-VAE + GBC showed significant improvements:

- **AUC-ROC:** The model achieved an average AUC-ROC score of 0.94, outperforming baseline models, which averaged between 0.87-0.89.
- **Precision and Recall:** For well-represented classes, the RNN-VAE + GBC surpassed baseline models by 5-10% in precision and recall. For rare classes like TDEs and CaRTs, the model improved precision and recall by 10-15% over the baselines.
- **Reduced Misclassifications:** The hybrid architecture allowed for better handling of complex decision boundaries, reducing misclassifications, particularly among supernova subclasses

Comparison with Baseline Models:
The RNN-VAE + GBC model was compared against baseline models (Figure 7) used in the original PLAsTiCC competition, such as Random Forests, Support Vector Machines (SVMs), and Convolutional Neural Networks (CNNs) (Pasquet et al., 2018). The RNN-VAE + GBC showed significant improvements:
AUC-ROC: The model achieved an average AUC-ROC score of 0.94, outperforming baseline models, which averaged between 0.87-0.89.

- **Precision and Recall:** For well-represented classes, the RNN-VAE + GBC surpassed baseline models by 5-10% in precision and recall. For rare classes like TDEs and CaRTs, the model improved precision and recall by 10-15% over the baselines.
- **Reduced Misclassifications:** The hybrid architecture allowed for better handling of complex decision boundaries, reducing misclassifications, particularly among supernova subclasses.

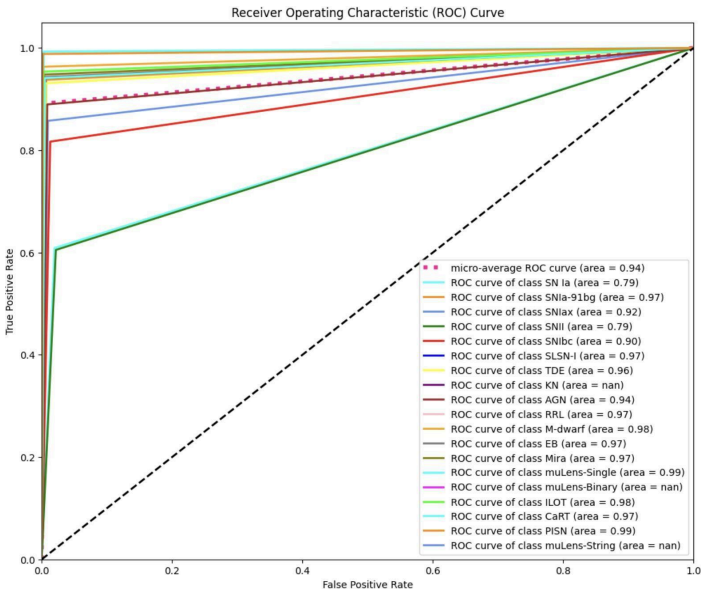

Figure 6: The Receiver Operating Characteristic (ROC) Curve shows the relationship between the false and true positive rate, demonstrating the discrimination ability of the model. This ROC curve demonstrates the micro-averages of each class. Note: created by student researcher.

Limitations:
Despite the promising results of the RNN-VAE + Gradient Boosting Classifier model, several limitations need to be addressed:

- **Overlapping Light Curve Characteristics:** One of the main challenges observed was the misclassification between transient classes with similar light curve features, such as Type Ia (SN Ia) and Type Iax supernovae. The model sometimes struggled to differentiate these classes due to their overlapping photometric features, particularly in the early stages of their light curves. This limitation indicates that additional distinguishing features, such as spectral data, could improve classification accuracy.
- **Handling Noisy Data:** While the RNN-VAE architecture includes mechanisms to handle incomplete data, it remains sensitive to noise and distortions in light curves. In cases where data quality was compromised (e.g., due to low signal-to-noise ratios or gaps in observations), the reconstruction of light curves retained some distortions, which affected classification accuracy. Future improvements could involve advanced denoising methods or integrating attention mechanisms to better focus on cleaner data segments.
- **Computational Complexity:** The combined RNN-VAE and Gradient Boosting Classifier architecture is computationally intensive, requiring significant processing power for training and inference. This presents challenges for scaling the model to larger datasets or deploying it in real-time settings. Optimizations like model pruning, quantization, or distributed computing could help address

this limitation, making the model more efficient without sacrificing accuracy.

- **Class Imbalance Effects:** Although the model effectively handled class imbalance through oversampling and class-weight adjustments, these techniques introduced a slight increase in false positives for certain classes with overlapping characteristics. This trade-off highlights the need for more sophisticated strategies, such as synthetic data generation through Generative Adversarial Networks (GANs), to balance recall and precision without increasing misclassifications.
- **Generalization to New Data:** While the model demonstrated strong performance on the PLAsTiCC dataset, its generalization to new, unseen datasets remains uncertain. Variations in telescope sensitivity, observational conditions, or transient event characteristics could impact model performance. Additional training with more diverse datasets could improve robustness and generalizability.

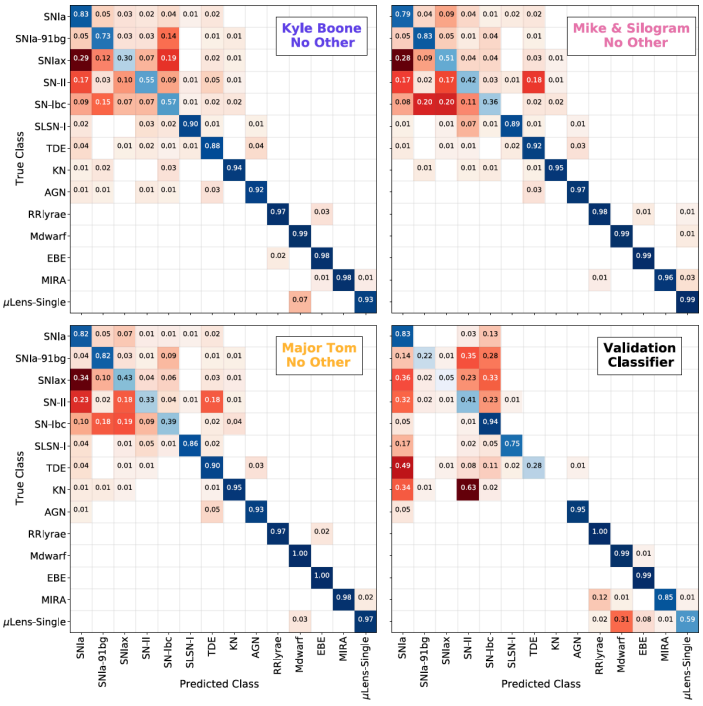

Figure 7: The confusion matrices showcase the pseudo-confusion matrices of the top three submissions of the Kaggle contest. Note: This figure is adapted from Hložek et al. 2020.

Future Work:

To address these limitations and further enhance the model's performance, several future directions are proposed:

- **Incorporating Spectral Data:** Integrating spectral data alongside photometric data could improve feature extraction, particularly for distinguishing between similar transient classes (Boone, 2019).
- **Attention Mechanisms:** Adding attention mechanisms to the RNN-VAE architecture could enhance feature learning by focusing on the most relevant temporal patterns in the light curves (Villar et al., 2019).
- **Advanced Augmentation:** Utilizing advanced data augmentation techniques, such as synthetic light curve generation via Generative Adversarial Networks (GANs), could help balance the dataset and provide more training samples for rare classes (Rezende et al., 2014). Hybrid Models: Exploring hybrid models that combine RNN-VAE with other algorithms, such as XGBoost or Random Forests, could offer complementary strengths in handling imbalanced datasets and improving decision boundaries.
- **Real-Time Optimization:** Developing a lightweight version of the model, using techniques like pruning or quantization, could reduce computational demands and enable faster inference for real-time applications (Pasquet et al., 2018).
- **Ensemble Variations:** Investigating ensemble variations like stacking or bagging could improve classification accuracy and robustness, especially for classes with overlapping features. By pursuing these directions, the model can be further refined to enhance its classification accuracy, robustness, and adaptability to the evolving demands of real-time astronomical surveys.

With Thanks & Acknowledgement

Portions of this work were conducted in the Minnesota Nano Center, which is supported by the National Science Foundation through the National Nano Coordinated Infrastructure Network (NNCI) under Award Number ECCS-1542202.

I dedicate my work in this study to Henrietta Lacks. Mrs. Lacks died not knowing how greatly she would impact science; it is with my sincere gratitude that I was given permission to use HeLa cells.* I want to express a multitude of thanks to Dr. James Marti for his co-authorship, advising, and generosity in inviting me to use the Bio-Nano Lab, and to Matthew Johnson, who offered excellent guidance, mentorship, and training. Lastly, it is because of Dr. Kristyn VanderWaal Mills that I was introduced to this amazing opportunity – her support throughout this study made my work possible.

*The isolation and immortalization of HeLa cells changed the scientific landscape in the 20th century. HeLa cells, named for Henrietta Lacks, were first collected by two Johns Hopkins researchers in February, 1951, and through the efforts of later researchers, were established as a standard for cell laboratory propagation and studies [14]. The original HeLa cells were obtained from 31-year-old patient Henrietta Lacks, who was being treated for cervical cancer at Johns Hopkins Hospital [10]. The cells from Henrietta Lacks were obtained without her knowledge, as permission was not required for cell harvesting in 1951. Henrietta Lacks later died from the cancer on October 4, 1951.

References

Boone, K. (2019). Avocado: Photometric classification of astronomical transients with Gaussian process augmentation. *The Astronomical Journal, 158*(6), 257. https://doi.org/10.3847/1538-3881/ab5182

Fitzpatrick, E. L. (1999). Correcting for the effects of interstellar extinction. *Publications of the Astronomical Society of the Pacific, 111*(755), 63–75. https://doi.org/10.1086/316293

Heymann, Y. (2012). Redshift adjustment to the distance modulus. *Progress in Physics, 8*.

Hložek, R., et al. (2020). Results of the photometric LSST astronomical time-series classification challenge (PLAsTiCC). *arXiv*. https://arxiv.org/abs/2012.12392

Ivezić, Ž., et al. (2019). LSST: From science drivers to reference design and anticipated data products. *The Astrophysical Journal, 873*(2), 111. https://doi.org/10.3847/1538-4357/ab042c

Kessler, R., et al. (2019). Models and simulations for the photometric LSST astronomical time series classification challenge (PLAsTiCC). *Publications of the Astronomical Society of the Pacific, 131*(1003), 094501. https://doi.org/10.1088/1538-3873/ab26f1

Kingma, D. P., & Welling, M. (2022). Auto-encoding variational Bayes. *arXiv*. https://arxiv.org/abs/1312.6114

Möller, A., et al. (2016). Photometric classification of type Ia supernovae in the SuperNova Legacy Survey with supervised learning. *Journal of Cosmology and Astroparticle Physics, 2016*(12), 008. https://doi.org/10.1088/1475-7516/2016/12/008

Morawski, F., et al. (2021). Anomaly detection in gravitational waves data using convolutional autoencoders. *Machine Learning: Science and Technology, 2*(4), 045014. https://doi.org/10.1088/2632-2153/abf3d0

Muthukrishna, D., et al. (2019). RAPID: Early classification of explosive transients using deep learning. *Publications of the Astronomical Society of the Pacific, 131*(1005), 118002. https://doi.org/10.1088/1538-3873/ab1609

Pasquet, J., et al. (2018). Photometric redshifts from SDSS images using a convolutional neural network. *Astronomy & Astrophysics, 621*, A26. https://doi.org/10.1051/0004-6361/201833617

Villar, V., et al. (2019). Supernova photometric classification pipelines trained on spectroscopically classified supernovae from the Pan-STARRS1 Medium-deep Survey. *The Astrophysical Journal, 884*, 83. https://doi.org/10.3847/1538-4357/ab418c

Villar, V. A., et al. (2021). A deep-learning approach for live anomaly detection of extragalactic transients. *The Astrophysical Journal Supplement Series, 255*(2), 24. https://doi.org/10.3847/1538-4365/ac0893