SunGuard: Solar Flare Classification using Deep Learning Techniques

Adnan Sawalha[1], Ihmaidan Al-Haj[1], Basil Alajlouni[1], Dunia Alatoom[1], Ahmad Almanasrah[2], Nikos Nikolaou[3], and Omar Al-Kadi[1]

[1]Department of Artificial Intelligence, University of Jordan, Amman 11942, Jordan
[2]Department of Computer Information Systems, University of Jordan, Amman 11942, Jordan
[3]Department of Physics and Astronomy, University College London, London WC1E 6BT, United Kingdom
{adn0221536, ahm2220516, bas0227843, dny0205996, ahm0227255, o.alkadi}@ju.edu.jo
n.nikolaou@ucl.ac.uk

Abstract

Severe ($\geq M$-class) solar flares pose significant risks to modern technological infrastructure, yet reliable classification remains challenging due to extreme class imbalance and the need to exploit localized magnetic and morphological precursors across heterogeneous solar observations. This paper introduces a modular multi-view framework for severe solar flare classification on the SDOBenchmark dataset, which provides ten synchronized Solar Dynamics Observatory channels. Each wavelength is modeled independently using a pretrained EfficientNetV2-S backbone to learn wavelength-specific representations. To adapt single-channel solar images to the three-channel input requirement, we construct a frequency–spatial tensor per image using

the original intensity, FFT magnitude, and FFT phase, followed by ImageNet normalization. The resulting per-wavelength embeddings are fused using two feature-level strategies, a 1D residual neural fusion model that learns nonlinear cross-wavelength interactions while preserving channel structure, and an XGBoost ensemble operating on concatenated embeddings as a strong classical baseline. To further mitigate the severe $\sim$1:15 class imbalance, we evaluate controlled minority-class augmentation using a per-wavelength WGAN-GP, and assess synthetic image realism using PSNR and SSIM in addition to classification metrics. Performance is evaluated with validation-driven threshold selection that maximizes the True Skill Statistic (TSS). Experiments show that XGBoost fusion provides the strongest real-data baseline (TSS 0.500 ± 0.016), while the residual fusion model benefits from moderate GAN augmentation (TSS improving from 0.412 ± 0.024 to ≈ 0.435 in intermediate synthetic-to-real ratios), highlighting the distinct augmentation sensitivity between neural and tree-based fusion approaches.

Keywords: Solar flare classification; Multi-wavelength fusion; EfficientNetV2; Residual neural fusion; Generative Adversarial Networks; XGBoost; Class imbalance

1 Introduction

Solar flares are among the most energetic and complex phenomena in astrophysics. They are sudden and intense bursts of electromagnetic radiation originating from the Sun's atmosphere, typically associated with the rapid release of magnetic energy stored in the solar corona (outer atmosphere). This energy release occurs primarily through a process known as *magnetic reconnection*, in which twisted magnetic field lines, as shown in Figure 1, suddenly realign and discharge energy in the form of radiation, heat, and accelerated particles. Solar flares radiate across the entire electromagnetic spectrum, from radio waves to gamma rays, with each type of radiation having a distinct impact on Earth's environment. Depending on their strength and characteristics, solar flares can significantly influence the space environment around Earth, sometimes causing disruptive and damaging effects on modern technological systems [1].

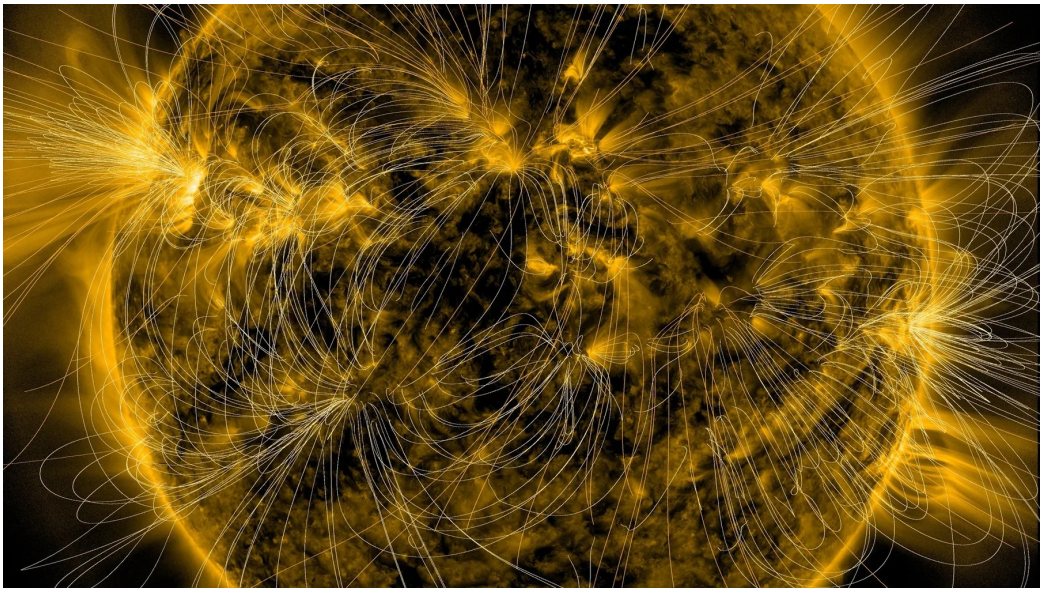

Figure 1: Sun's magnetic fields is overlaid on an image of the Sun captured in extreme ultraviolet light by NASA's Solar Dynamics Observatory. [1].

Solar flares are commonly classified according to the peak flux of X-rays they emit in the 1–8 Å range as measured by the Geostationary Operational Environmental Satellite (GOES). This classification system defines five categories: A, B, C, M, and X, with each successive letter representing a tenfold increase in intensity. A- and B-class flares typically reflect background solar activity and pose little risk to Earth, whereas C-class flares are weak but occasionally produce minor space weather disturbances. M-class flares are moderate events that can disrupt radio communication at high latitudes, while X-class flares, the most extreme category, are capable of triggering widespread communication blackouts, satellite malfunctions, and geomagnetic storms with cascading effects on power grids and navigation systems [2]. The classes are further refined with a linear scale from 1 to 9 (e.g., X1 to X9), although some exceptional flares exceed this limit [3].

Flares are often associated with *coronal mass ejections* (CMEs), which are massive eruptions of magnetized plasma expelled into interplanetary space. When directed toward Earth, CMEs can trigger geomagnetic storms that disturb Earth's magnetic field, leading to phenomena such as radio blackouts, power grid failures, and enhanced auroral displays. Although not every flare produces a CME, the two phenomena frequently coincide, amplifying their

collective impact on space weather hazards [4]. The disruptive consequences of solar flares and CMEs have been recognized throughout history. The Carrington Event of 1859 remains the most famous example, when an intense solar outburst observed by Richard Carrington coincided with widespread geomagnetic disturbances. Telegraph systems across Europe and North America malfunctioned, sparking fires in offices, while auroras were observed as far south as the Caribbean [5]. More recently, a geomagnetic storm induced by a powerful solar flare and CME collapsed the Hydro-Québec power grid in Canada, leaving six million people without electricity for nearly nine hours [6]. Another notable episode was the "Halloween Storms", which included an X28 flare-the largest ever recorded by GOES instruments-disrupting satellite operations, aviation, and polar communication routes [3]. In July 2012, Earth narrowly avoided a Carrington-scale CME; simulations suggest that a direct impact could have caused global technological disruptions comparable to or worse than the 1859 event [7]. Modern society's reliance on technology makes it increasingly vulnerable to the effects of solar activity. Satellites are susceptible to energetic particles that can damage electronics, degrade solar panels, or alter orbital trajectories [1]. Aviation, particularly on polar routes, suffers from high-frequency communication blackouts and increased radiation exposure [8]. Navigation systems such as Global Positioning System (GPS) often experience errors or outages, disrupting aviation, shipping, and precision agriculture [9]. Power grids are at risk from geomagnetically induced currents (GICs), which can overload transformers and cause widespread outages [2]. Even military operations dependent on secure communication and navigation are affected [10].

Severe ($\geq$M/X) solar flares are strongly related to the magnetic complexity of active regions, especially the presence of strong-gradient polarity inversion lines (PILs) and highly sheared magnetic configurations. These structures concentrate large amounts of free magnetic energy and helicity [11]. Observational studies show that major flares almost never occur without a strong-gradient PIL located inside the strong-field core of an active region [11]. In addition, regions that frequently produce flares and coronal mass ejections often display clear morphological features such as δ-spots, filaments, and sigmoidal structures, which are formed along strongly sheared PILs, as illustrated in Figure 2 [12]. Together, these observations indicate that flare-related information is highly localized near narrow, high-contrast magnetic boundaries and organized across multiple spatial scales, ranging from compact magnetic flux regions to extended structures aligned with PILs. For this

reason, effective flare classification models must preserve the two-dimensional spatial structure of the data rather than relying only on aggregated scalar parameters, which can remove important information about the magnetic field configuration [13].

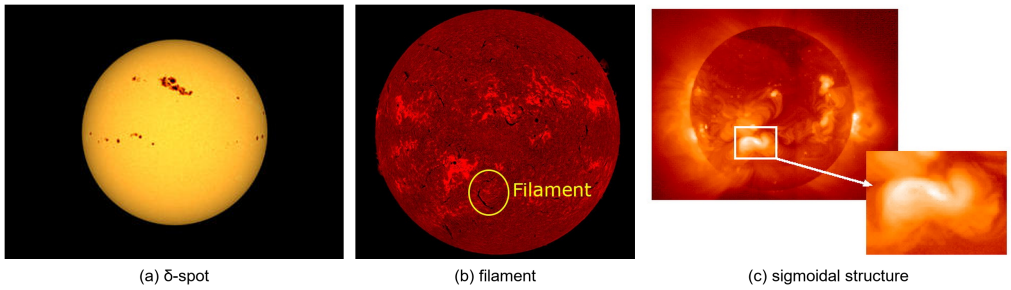

Figure 2: Examples of common morphological features observed in flare-productive solar active regions. (a) A δ-spot configuration showing closely packed opposite magnetic polarities, (b) a solar filament tracing a polarity inversion line, and (c) a sigmoidal coronal structure indicating a strongly sheared magnetic configuration.

Based on these physical considerations, convolutional neural networks are used, as they are well suited to extract spatial features directly from solar images. Strong-gradient PILs appear as thin, high-contrast boundaries that can be naturally captured by convolutional filters operating on local neighborhoods. As the network depth increases, information from local features is progressively integrated, allowing the model to represent larger-scale magnetic structures such as long PILs and sigmoidal patterns. EfficientNet is selected as the backbone for each wavelength because its balanced scaling of depth, width, and resolution helps preserve fine spatial details without excessive downsampling [14]. This is particularly important for resolving narrow PIL-related structures across wavelengths that probe different temperatures and layers of the solar atmosphere. After feature extraction is performed separately for each wavelength, information from multiple channels is combined at the feature level instead of the image level. Residual networks are used to integrate encoded representations from different wavelengths, as residual connections allow deeper feature transformations without causing optimization problems [15]. This design enables the fusion model to capture interactions between learned features that reflect extended, multi-scale magnetic structures

linked to strongly sheared configurations, while operating on compact feature representations rather than high-dimensional images. In parallel, XGBoost is used as an alternative feature-level fusion method. Gradient-boosted decision trees provide a regularized and sparsity-aware framework that works well with structured feature input [16]. Since different wavelengths represent complementary physical processes at different temperatures and atmospheric heights, their influence on flare initiation is inherently nonlinear and dependent on feature interactions. Tree-based boosting naturally captures such relationships by learning decision rules that model joint dependencies among wavelength-specific predictors, making XGBoost an effective approach for combining heterogeneous features into a final flare prediction.

Building on these design choices, this work introduces a modular, multi-view framework for severe solar flare classification, with the following main contributions:

a) A multi-stage, wavelength-specific learning framework that captures physically meaningful representations from ten synchronized Solar Dynamics Observatory (SDO) channels.

b) Two embedding-based fusion strategies employing ensemble models, evaluated on the SDOBenchmark dataset.

c) The use of GAN-based data augmentation to mitigate the severe $\sim$1:15 class imbalance.

The remainder of this paper is organized as follows: Section 2 reviews *related work* in solar flare classification and forecasting using statistical, machine learning, and deep learning approaches. Section 3 details the *Methodology*, including dataset description, preprocessing, wavelength-specific representation learning, fusion strategies, imbalance handling, and threshold optimization. Section 4 presents the *Results*, followed by Section 5 which is a comparative analysis and discussion, ending with the conclusion in Section 6.

2 Literature Review

Solar flare classification and forecasting has progressed through several methodological paradigms, evolving from statistical baselines and engineered magnetic parameters to image-based deep learning and, more recently, attention-driven

architectures. Despite steady improvements in predictive skill, persistent challenges remain, particularly severe class imbalance, limited interpretability of multi-modal fusion, and difficulty disentangling the contribution of heterogeneous solar observations. Existing approaches can be broadly categorized into four methodological subtypes: (i) statistical and feature-based models, (ii) image-based convolutional models, (iii) hybrid and fusion-based systems, and (iv) attention-based and Transformer-driven models. This categorization reflects increasing model capacity and representational flexibility, while also highlighting unresolved limitations that motivate the design choices in this work.

Early solar flare forecasting relied on statistical formulations such as Bayesian inference and Poisson-based event modeling [17]. These approaches provided transparent probabilistic baselines but were limited by their reliance on historical flare rates and their inability to encode spatial information related to active-region morphology. The introduction of evaluation metrics such as the True Skill Statistic (TSS) improved benchmarking under class imbalance conditions [18,19], yet predictive gains remained modest. With the availability of SDO/HMI data, classical machine learning methods trained on engineered magnetic field parameters, such as SHARP features, became prominent. Bobra and Couvidat [20] demonstrated that support vector machines trained on these features could achieve operationally relevant performance. Subsequent studies showed that forecast quality depended more on feature selection, data partitioning, and evaluation protocol than on the specific classifier employed [21, 22]. While these methods offer interpretability, they remain constrained by hand-crafted inputs and limited capacity to model complex spatial structures.

The adoption of convolutional neural networks (CNNs) enabled direct learning from magnetogram and EUV images, allowing models to capture spatial patterns associated with flare productivity [23]. CNN-based approaches consistently outperformed feature-driven methods by eliminating manual feature engineering. However, most implementations relied on cropped active-region patches or single-time snapshots, limiting their ability to capture global solar context and long-range spatial interactions [24]. Several studies proposed hybrid CNN-based pipelines to improve reliability and reduce false alarms [25], while others incorporated simplified physical constraints to enhance robustness [26]. Nevertheless, CNN architectures remain fundamentally limited by fixed receptive fields, sensitivity to class imbalance, and difficulty scaling to heterogeneous multi-wavelength inputs without tightly coupled architectures.

To leverage complementary information from multiple data sources, several works introduced hybrid systems that combine deep feature extractors with classical classifiers or sequential models. Operational frameworks such as Deep Flare Net [27] demonstrated near-real-time forecasting capability but relied heavily on engineered inputs. More complex fusion pipelines integrated CNNs, dense networks, and recurrent models to jointly capture spatial and temporal dependencies [28]. While these systems often achieved improved predictive performance, they typically employed tightly coupled end-to-end architectures. This design makes it difficult to isolate the contribution of individual observation channels, systematically address class imbalance at different stages, or conduct controlled ablation studies. These limitations reduce interpretability and hinder principled analysis of multi-modal solar data.

Attention mechanisms were introduced to overcome the locality constraints of CNNs by enabling global dependency modeling. Transformer architectures applied to time-series magnetic parameters improved temporal modeling but continued to rely on engineered inputs [29]. Vision Transformers (ViTs) extended self-attention to image-based inputs, enabling global spatial reasoning and improved performance on magnetogram data [30]. More recently, large-scale Transformer-based foundation models for heliophysics, such as Surya [31], have demonstrated the potential of attention-driven learning across diverse solar datasets. However, such models are computationally expensive, difficult to deploy operationally, and often trained in end-to-end configurations that limit interpretability and systematic analysis of individual wavelengths. Moreover, the integration of multi-wavelength data is typically implicit, obscuring the physical contribution of each observation channel.

Despite substantial progress, three key limitations persist across existing solar flare forecasting approaches. First, severe class imbalance is often handled through simplistic resampling strategies or threshold-agnostic objectives, limiting sensitivity to rare but operationally critical events. Second, many models rely on tightly coupled architectures that hinder interpretability and prevent controlled evaluation of individual data sources. Third, multi-wavelength observations are frequently fused in an opaque manner, obscuring their physical contribution to classification performance. These limitations motivate the modular, multi-stage framework proposed in this work. By learning wavelength-specific representations independently, integrating them through embedding-based fusion, and applying validation-driven threshold optimization, the proposed approach enables systematic analysis of multi-

wavelength solar observations while directly addressing class imbalance and interpretability concerns.

3 Methodology

3.1 SDOBenchmark Dataset

NASA's Solar Dynamics Observatory (SDO) is a space-based mission designed to continuously monitor the Sun's atmosphere, magnetic fields, and radiative outputs at high spatial and temporal resolution [32]. Its instruments provide comprehensive, full-disk observations across multiple wavelengths and magnetic diagnostics, enabling detailed analysis of solar activity across different atmospheric layers. Among these instruments, the Atmospheric Imaging Assembly (AIA) [33] and the Helioseismic and Magnetic Imager (HMI) [34] are particularly relevant for solar flare classification, as they capture complementary thermal and magnetic information critical to flare initiation and evolution. Building on SDO observations, the SDOBenchmark dataset [35] has emerged as a widely used benchmark for solar flare classification and prediction. It provides synchronized, multi-modal samples curated specifically for machine learning research and reproducible model evaluation. Each sample includes ten temporally aligned images: eight EUV channels from AIA and two magnetic field measurements from HMI.

Flare labels are derived from GOES X-ray flux measurements, with M- and X-class flares labeled as *severe* and A-, B-, and C-class flares labeled as *non-severe*. Formally, the peak soft X-ray flux measured in the GOES 1–8 Å band over a predefined temporal window Δt is defined as

$$F_{\mathrm{GOES}}^{\mathrm{peak}} = \max_{t \in \Delta t} F_{1\text{–}8\ \text{Å}}(t), \qquad \text{Severe} \iff F_{\mathrm{GOES}}^{\mathrm{peak}} \geq 10^{-5}\ \mathrm{W, m^{-2}}. \quad (1)$$

The dataset is formulated as a binary classification task. This standardized labeling and temporal setup facilitates fair comparison across algorithms while emphasizing two core challenges in operational flare classification: extreme class imbalance and the need to extract meaningful precursors from high-dimensional solar image data. Although the dataset provides synchronized multi-wavelength observations, each channel encodes different physical information. This characteristic supports modeling approaches that learn wavelength-specific representations before integrating information across modalities. Among the SDO instruments, AIA observes the full solar disk

across multiple EUV channels, each sensitive to plasma at different temperatures and heights in the solar atmosphere. For example, the 94 Å and 131 Å channels capture extremely hot plasma ($\sim 10^6$–10^7 K) typically associated with flare initiation and impulsive energy release, whereas the 171 Å and 193 Å channels emphasize quieter coronal loops and evolving active regions. Other channels, including 211 Å, 304 Å, and 335 Å, provide complementary views of the hotter corona and chromospheric emission, while the 1700 Å continuum images the upper photosphere and sunspot structure.

Complementing these observations, HMI magnetograms map the photospheric magnetic field, providing crucial information on magnetic complexity, polarity inversion lines, and energy storage; key precursors to flare activity. In physical terms, the energy available to power eruptive events is commonly expressed as the magnetic free energy stored in non-potential field configurations,

$$E_{\text{free}} \approx \int_V \frac{B^2 - B_{\text{pot}}^2}{2\mu_0}, dV; ; \geq 0, \qquad B_{\text{pot}} = \arg\min \int_V B^2, dV. \tag{2}$$

Regions exhibiting strong field gradients and complex polarity structures tend to store larger amounts of free magnetic energy, increasing their likelihood of producing energetic flares. By combining these ten synchronized channels (Figure 3), researchers obtain a multi-layered, multi-thermal view of solar activity. This property is central to the multi-view learning strategy adopted in this work, in which wavelength-specific representations are learned independently and later fused to capture cross-layer interactions relevant to severe flare classification.

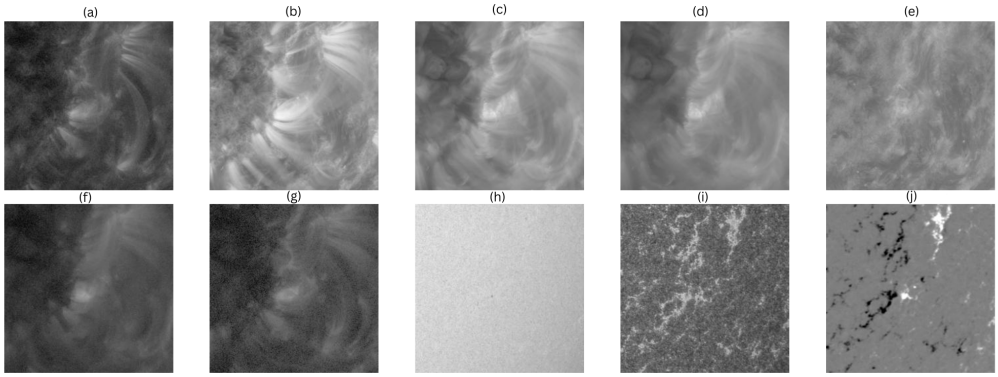

Figure 3: Sample of the ten synchronized image channels provided in the SDOBenchmark dataset. Panels (a)–(j) correspond to the following channels: (a) 131 Å, (b) 171 Å, (c) 193 Å, (d) 211 Å, (e) 304 Å, (f) 335 Å, (g) 94 Å, (h) HMI continuum intensity, (i) 1700 Å, and (j) HMI line-of-sight magnetogram.

3.2 Data Preprocessing and Augmentation

All input observations are processed using a unified preprocessing pipeline to ensure consistency across different wavelengths and to support stable model training. Each solar observation is represented as a single-channel image corresponding to a specific wavelength or magnetogram measurement. Before being fed into the models, all images are resized to a fixed spatial resolution of 128×128 pixels. This resolution is chosen to preserve important solar structures while keeping computational costs manageable during large-scale experiments. Pixel intensity values are converted to floating-point format and because the pre-trained EfficientNetV2 model expects three-channel RGB inputs while solar observations are single-channel grayscale images, instead of replicating them across three channels, a Fourier Frequency Transformation (FFT) is used to generate a three-channel tensor, consisting of the original image, the FFT magnitude, and the phase spectrum, as illustrated in figure 4 and formally described in algorithm 1.

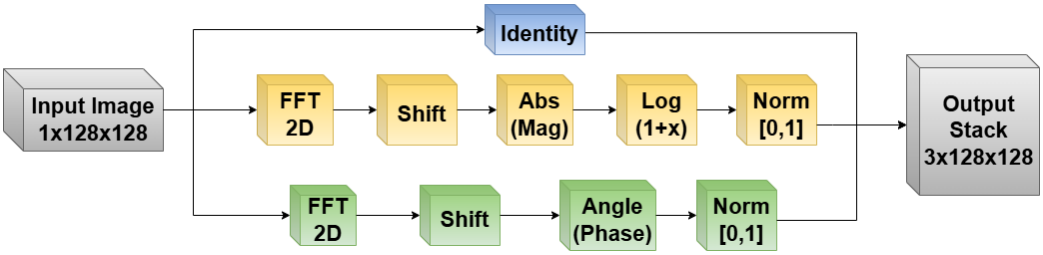

Figure 4: Frequency-domain feature extraction pipeline using Fourier Frequency Transformation.

Algorithm 1 Frequency–Spatial Feature Construction from a Single Wavelength Image

Require: Grayscale image I
Ensure: Three-channel feature tensor $\mathbf{T}$

$\text{Identity} \leftarrow I/255.0$
$\mathbf{F} \leftarrow \text{FFT2D}(I)$
$\mathbf{F}_s \leftarrow \text{FFTShift}(\mathbf{F})$
$\text{Mag} \leftarrow \log(1 + |\mathbf{F}_s|)$
$\text{Mag} \leftarrow \dfrac{\text{Mag} - \min(\text{Mag})}{\max(\text{Mag}) - \min(\text{Mag})}$
$\text{Phase} \leftarrow \angle(\mathbf{F}_s)$
$\text{Phase} \leftarrow \dfrac{\text{Phase} - \min(\text{Phase})}{\max(\text{Phase}) - \min(\text{Phase})}$
$\mathbf{T} \leftarrow \text{Stack}([\text{Identity}, \text{Mag}, \text{Phase}], \text{axis} = 0)$
return $\mathbf{T}$

After this transformation, the resulting three-channel tensor is normalized using the standard ImageNet mean and standard deviation ($\mu = [0.485, 0.456, 0.406]$, $\sigma = [0.229, 0.224, 0.225]$) to match the input scaling of the pretrained EfficientNetV2-S backbone.

To improve generalization and limit overfitting, a set of data augmentation techniques that are physically plausible is applied during training. These augmentations are chosen to increase data diversity while keeping the magnetic and morphological structures related to solar activity unchanged. Specifically, random horizontal and vertical flips are applied to input images with equal probability. These transformations are physically valid for solar disk observations and allow the model to see different spatial arrangements of active

regions. In addition, mild photometric changes are introduced, including small-scale intensity jitter and additive Gaussian noise. These perturbations simulate differences in instrument response, photon noise, and observing conditions, helping the model become more robust to minor intensity changes. All augmentations are applied only to the training data, while validation and test images remain unchanged to ensure a fair and unbiased evaluation of model performance.

Solar flare classification is characterized by severe class imbalance, with high-energy flare events occurring far less frequently than non-flaring or low-intensity events. To address this issue without using synthetic data at this stage, a class-balanced sampling strategy is applied during training. In this approach, samples from the minority class are selected with a higher probability than their natural occurrence in the dataset. This effectively increases the presence of severe flare events during training while still using only real data. This approach reduces bias toward the majority class and improves sensitivity to rare but operationally critical flare events. Along with sampling adjustments, the training objective is designed to place greater emphasis on difficult and minority-class examples. The loss function assigns higher penalties to misclassified severe events while reducing the influence of easily classified majority samples, and is formulated using a focal loss function,

$$\mathcal{L}_{\text{focal}}(p, y) = -, \alpha, y, (1-p)^{\gamma} \log p; -; (1-\alpha), (1-y), p^{\gamma} \log(1-p), \quad y \in 0, 1. \tag{3}$$

Here, p denotes the predicted probability of the severe class, α controls the relative weighting of minority and majority classes, and γ modulates the down-weighting of well-classified examples. This formulation encourages the model to focus on challenging cases and prevents convergence toward trivial solutions dominated by the majority class. All preprocessing and non-GAN augmentation steps are applied independently to each wavelength-specific dataset. This ensures that every single-wavelength classifier is trained on a consistent and balanced input distribution. The use of synthetic data generated by generative models is addressed separately and discussed in Section 3.6.

3.3 Proposed System Architecture

The proposed system follows a modular, multi-stage architecture designed to use complementary information from multiple solar observation wavelengths while maintaining scalability and interpretability, illustrated in Figure 5. Instead of training a single large model on multi-channel inputs, the framework separates the classification task into two main parts: learning representations independently for each wavelength and then combining them through multi-wavelength fusion. This design allows each component to be optimized separately and makes it possible to clearly analyze the contribution of individual wavelengths as well as their combined effect. The pipeline consists of four sequential stages: (i) wavelength-specific data preparation and augmentation, (ii) single-wavelength feature learning, (iii) multi-wavelength fusion, and (iv) prediction thresholding and evaluation. Each stage operates independently and produces well-defined intermediate outputs that serve as inputs to subsequent stages.

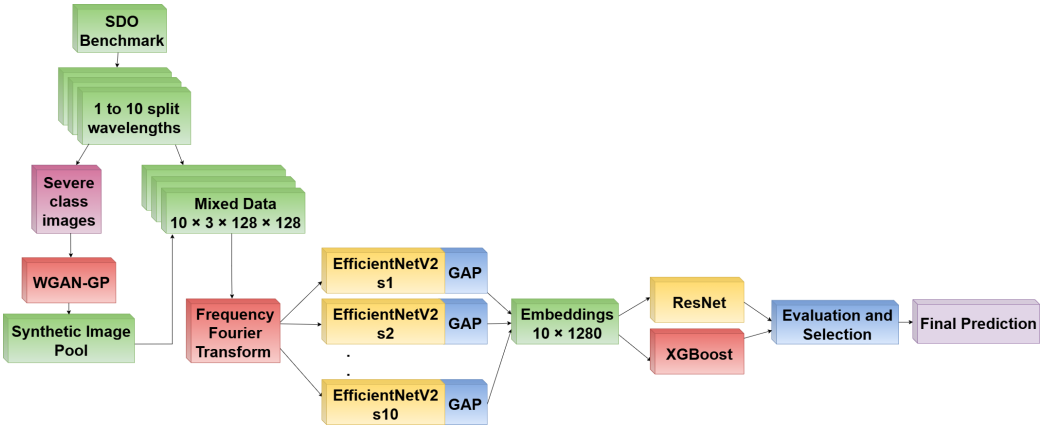

Figure 5: Overview of the proposed end-to-end framework

In the first stage, each wavelength is processed separately using the preprocessing and augmentation procedures described in Section 3.2. When generative augmentation is used, synthetic samples produced by generative adversarial networks are added at this stage using controlled ratios between synthetic and real data. These ratios are varied across experiments to study how the amount of synthetic data affects later model performance. In the second stage, a dedicated classifier is trained for each wavelength independently. These models are responsible for learning wavelength-specific representations

that capture spatial patterns and intensity variations related to solar flare activity. In addition to generating probability estimates, the models are used to extract compact feature embeddings that summarize the learned information for each observation. These embeddings provide a consistent representation that can be reused in later stages. The third stage performs multi-wavelength fusion by combining the embeddings produced by the individual wavelength-specific models. Fusion is performed using two different approaches: a 1D residual neural network–based fusion model, which is a deep neural network that learns nonlinear relationships across wavelengths, and XGBoost, a classical machine-learning model that operates on the same learned embeddings. Using both approaches makes it possible to compare deep and non-deep fusion methods while keeping the learned feature representations fixed, which helps isolate the effect of the fusion strategy itself. Finally, the predicted probabilities are transformed into binary flare predictions using a dynamic thresholding approach based on validation data. Instead of using a fixed decision threshold, an optimal threshold is selected on the validation set by maximizing the TSS. Once the threshold is determined, it is kept fixed and applied directly to the held-out test set. This ensures that the final evaluation remains fair and unbiased, and that test results are not influenced by information from the test data itself.

Generative data augmentation is included in the architecture as an optional component rather than a required one. When enabled, generative models are trained separately for each wavelength using only training data. The synthetic samples are then merged with real observations before training the wavelength-specific classifiers. Importantly, generative augmentation affects only the representation learning stage and does not directly influence fusion or evaluation. By varying the ratio of synthetic to real samples across experiments, the framework allows a systematic study of how generative augmentation impacts feature learning, fusion effectiveness, and sensitivity to rare severe events. This architectural design is guided by several key considerations. First, independent wavelength-specific learning allows each model to learn physically meaningful patterns specific to each observation type. Second, embedding-based fusion reduces computational complexity and enables flexible experimentation with different fusion methods without retraining the entire pipeline. Third, the separation of generative augmentation from fusion ensures that any performance improvements can be attributed to better feature representations rather than side effects of the fusion process.

3.4 Single-Wavelength Classifiers

To capture the different morphological and physical features that appear in various layers of the solar atmosphere, independent convolutional neural networks are trained for ten distinct observation channels. These include eight atmospheric wavelengths measured in Angstroms (94, 131, 171, 193, 211, 304, 335, and 1700), along with the HMI Continuum and HMI Magnetogram data. Training a separate model for each channel allows the system to learn features that are specific to the physical information provided by each type of observation. All single-wavelength classifiers use a pre-trained EfficientNetV2-S architecture as their backbone. This model is chosen because it offers high parameter efficiency and faster training compared to earlier EfficientNet versions and other classifiers. EfficientNetV2-S makes use of fused MBConv layers, which improve training stability and reduce memory access costs, making it suitable for large-scale solar image experiments.

The architecture employs EfficientNetV2-S with pretrained ImageNet weights, using fused-MBConv blocks, and compound scaling. The architecture of the EfficientNetV2-S model is shown in Figure 6. Each wavelength-specific model is trained independently to predict the probability of a solar flare event. The original classification head of the network is removed and replaced with a custom binary head that includes a dropout layer, with a dropout rate of 0.5 to reduce overfitting, followed by a linear layer that outputs a single prediction value. The SDOBenchmark dataset consists of training and test sets, but a validation set was generated by applying a 90/10 split on the training set in a stratified fashion to preserve the severe-to-non-severe ratio. Augmentation is applied to the training set by adding synthetic severe-class images, with the number of synthetic images being a proportion of the real severe images. The augmentation percentage varies for each experiment, while the validation and test sets remain real-only to avoid bias. The model is optimized using the AdamW optimizer with a learning rate of 5×10^{-5} and weight decay of 0.05. The training process runs for up to 24 epochs on a dedicated GPU, saving the checkpoints every two epochs. The model is evaluated using a dynamic threshold, which is selected to maximize the TSS on the validation set. This threshold is then applied to the test set for final evaluation. Once trained, these models function as feature extractors for the fusion stage. Latent representations are extracted from the final pooling layer of each network, producing a dense embedding vector $v_\lambda \in \mathbb{R}^{1280}$for each wavelength λ. These embeddings encode high-level semantic information related to active region

complexity and are used as inputs for the multi-wavelength fusion models. The implementation and experiment scripts are publicly available on GitHub.[1]

Each wavelength-specific EfficientNetV2-S model is trained first on its corresponding single-channel dataset. After training, the best checkpoint for each wavelength is frozen and used as a fixed feature extractor: embeddings are taken from the final pooling layer ($v_\lambda \in \mathbb{R}^{1280}$) for all samples, and only these embeddings are used to train the downstream fusion models.

3.5 Fusion Models

To combine information from different layers of the solar atmosphere and make use of the complementary nature of multiple wavelengths, the system applies a multi-stage fusion strategy. The input to this stage consists of the latent feature embeddings produced by the ten frozen single-wavelength classifiers. Instead of retraining the backbone networks, the fusion models operate only on these extracted representations. This design keeps the fusion process computationally efficient and allows the models to focus specifically on learning relationships between wavelengths. Two different fusion approaches are used: a deep learning–based model and a classical machine-learning ensemble.

The fusion approach is based on a Residual Network adapted for one-dimensional feature sequences. A diagram of the ResNet model pipeline used can be seen in Figure 7, right before Algorithm 2. Unlike simple fusion methods that immediately flatten all inputs, this architecture initially preserves the separation between wavelength channels. The ten embedding vectors are stacked to form a tensor of shape $(B, 10, 1280)$, where B is the batch size, 10 corresponds to the number of wavelength channels, and 1280 is the embedding dimension. The network starts with a one-dimensional convolutional layer that expands the input from 10 channels to a higher-dimensional feature space of 1024 channels. This is followed by a series of residual blocks that progressively transform the features from 1024 to 512, 256, and finally 128 channels. These blocks use one-dimensional convolutions with a kernel size of 3, combined with batch normalization and ReLU activations. This structure enables the model to learn complex and non-linear dependencies between features originating from different atmospheric layers. The network ends with a global adaptive average pooling layer that aggregates the learned features,

[1] `https://github.com/sunguard2026/SunGuard-Project/tree/main`

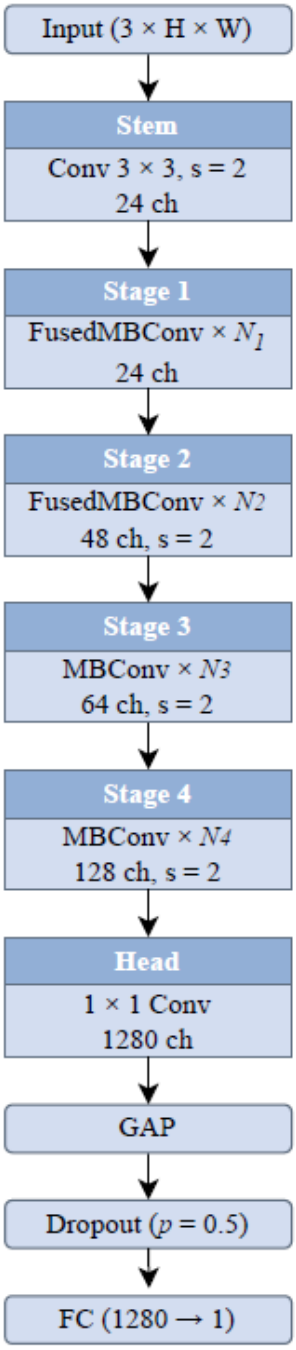

Figure 6: Architecture of the EfficientNetV2-S model.

followed by a final linear layer that outputs the predicted flare probability.

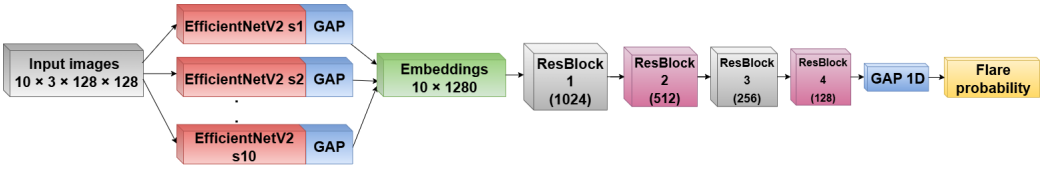

Figure 7: Residual neural fusion architecture for multi-wavelength solar flare classification.

As an alternative fusion method, an Extreme Gradient Boosting (XGBoost) classifier is used to provide a strong classical baseline. A general architecture of XGBoost is illustrated in figure 8. For this approach, the ten embedding vectors are concatenated into a single flattened feature vector of dimension 12,800. The XGBoost model is configured with 500 estimators and a learning rate of 0.05 to allow gradual and stable learning. To reduce overfitting in this high-dimensional feature space, the maximum tree depth is limited to 10, and both row and feature subsampling ratios are set to 0.8. These settings encourage the model to focus on the most informative features across wavelengths rather than fitting noise in the training data. Both fusion models are evaluated using the same validation-driven dynamic thresholding strategy based on the TSS. Using a shared threshold selection procedure ensures a fair and consistent comparison between the deep learning–based fusion model and the classical gradient boosting approach.

Algorithm 2 Residual Neural Fusion for Multi-Wavelength Solar Flare Classification

Require: Synchronized multi-wavelength observation $\mathcal{I} = \{I_\lambda\}_{\lambda=1}^{10}$
Ensure: Solar flare probability $\hat{y}$

Function ExtractSpatialFeatures($\mathcal{I}$)
  $\mathcal{F} \leftarrow \emptyset$
  $\mathcal{B} \leftarrow$ Truncated EfficientNetV2-S backbone
for each wavelength channel $I_\lambda \in \mathcal{I}$ **do**
    $I_\lambda^{(3)} \leftarrow \text{ChannelTransform}(I_\lambda)$
    $\mathbf{S} \leftarrow \mathcal{B}(I_\lambda^{(3)})$
    $\mathbf{z}_\lambda \leftarrow \text{GlobalAveragePooling}(\mathbf{S})$
    Append $\mathbf{z}_\lambda$ to $\mathcal{F}$
end for
  return Stack($\mathcal{F}$)
End Function

Function FusionWideResNet($\mathbf{X}$)
  $\mathbf{X} \leftarrow \text{ResidualBlock1D}(\mathbf{X}, 1024)$
  $\mathbf{X} \leftarrow \text{ResidualBlock1D}(\mathbf{X}, 512)$
  $\mathbf{X} \leftarrow \text{ResidualBlock1D}(\mathbf{X}, 256)$
  $\mathbf{X} \leftarrow \text{ResidualBlock1D}(\mathbf{X}, 128)$
  $\mathbf{p} \leftarrow \text{GlobalAveragePooling1D}(\mathbf{X})$
  $\hat{y} \leftarrow \sigma(\text{Linear}(\mathbf{p}))$
  return $\hat{y}$
End Function

$\mathbf{E} \leftarrow$ ExtractSpatialFeatures($\mathcal{I}$)
$\hat{y} \leftarrow$ FusionWideResNet($\mathbf{E}$)
return $\hat{y}$

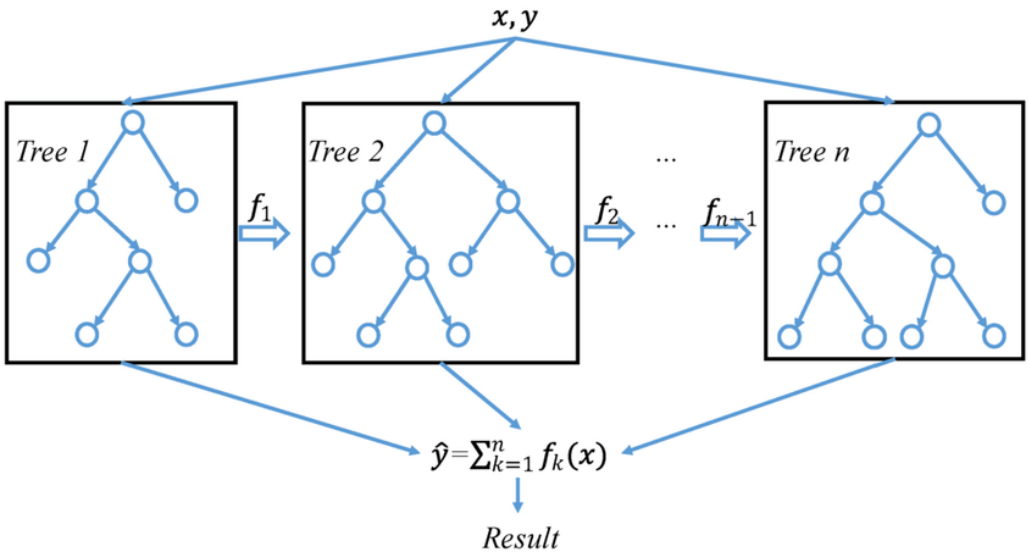

Figure 8: A general architecture of XGBoost, adapted from Wang et al. [36]

3.6 GAN-Based Data Augmentation

Generative Adversarial Networks (GANs) are deep generative models that generate realistic data samples through an adversarial training process between a generator and a discriminator. While standard GAN formulations have demonstrated strong generative capacity, they are often affected by training instability and mode collapse, particularly when applied to high-dimensional image data and severely imbalanced datasets. In initial experiments, multiple GAN variants were evaluated for solar image augmentation, including Vanilla GANs, DCGANs, and Wasserstein-based models. Based on training stability as well as the visual and structural quality of the generated images, the Wasserstein GAN with Gradient Penalty (WGAN-GP) was selected as it provides more stable convergence and better preservation of physically meaningful structures.

The GAN is trained separately for each wavelength channel and is applied only to augment the minority class, while the non-severe class is left unchanged. This design ensures that the augmentation process focuses solely on improving the representation of rare severe flare events. An illustration of the training flow of the WGAN-GP architecture is shown in figure 9.

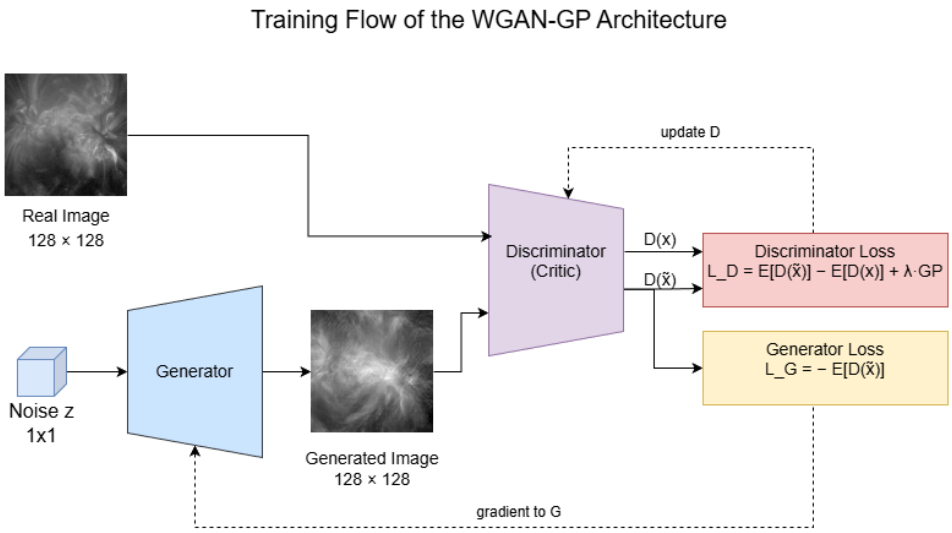

Figure 9: Training workflow of the WGAN-GP architecture used for synthetic solar image generation.

The data augmentation is performed at the single-wavelength level before multi-wavelength fusion. For a given wavelength channel, the WGAN-GP is trained using only real severe-class images from that channel. After training, the generator is used to create synthetic severe flare images, which are then combined with the original data at predefined augmentation ratios. This approach allows controlled analysis of how synthetic data affects classification performance. Before GAN training, the input images for the selected wavelength are preprocessed to ensure consistency. They are first converted to grayscale when applicable (intensity-only channels). And then all images are resized to a fixed resolution of 128 × 128 pixels using bicubic interpolation to ensure uniform input dimensions. The images are then converted to tensors and normalized to the range [-1, 1], matching the output range of the generator's final activation function. Although after generation, synthetic samples are inverse-normalized from the range $[-1, 1]$ back to standard intensity space and then passed through the same preprocessing pipeline as real images (resize → FFT channel construction in Algorithm 1 → ImageNet normalization) before training the single-wavelength classifiers. The GAN training dataset consists exclusively of severe-class images organized in a simple directory structure, and no non-severe images are included. This ensures that the generator learns only the distribution of severe solar flare

events for the specific wavelength.

A figure showcasing the flow of the generator training flow can be seen in figure 10. The generator is designed as a ResNet-style upsampling network that maps a latent noise vector $z \in \mathbb{R}^{128}$, sampled from a standard normal distribution, to a synthetic solar image of size 128×128 pixels. The architecture starts with a transposed convolution layer that projects the latent vector into a low-resolution feature map. This is followed by a sequence of residual upsampling blocks that gradually increase the spatial resolution through multiple stages (e.g. $4 \rightarrow 8 \rightarrow 16 \rightarrow 32 \rightarrow 64 \rightarrow 128$) until the target image size is reached. To better capture long-range spatial dependencies and large-scale structures within solar active regions, a self-attention layer is introduced at the 32×32 feature resolution. The final output layer applies a transposed convolution followed by a tanh activation function, producing images normalized to the range [-1, 1].

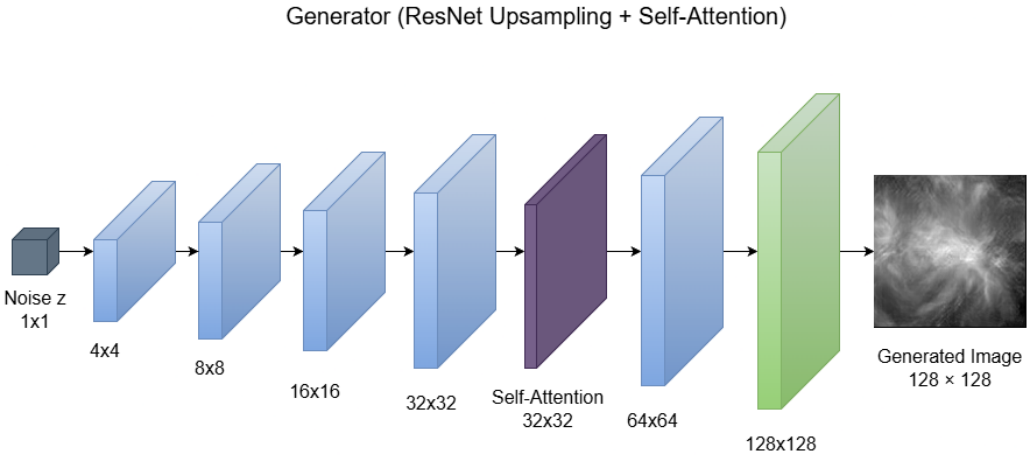

Figure 10: Training workflow of the WGAN-GP generator architecture.

The discriminator, referred to as the critic in the WGAN framework, is visualized in figure 11. It's implemented as a fully convolutional network that takes a 128×128 image as input and outputs a single scalar value without using a sigmoid activation. The critic consists of several strided convolutional layers with LeakyReLU activations, which progressively downsample the input image until a 1×1 spatial representation is obtained. The resulting scalar score reflects the critic's assessment of image realism and is used to approximate the Wasserstein distance between real and generated data

distributions when combined with the gradient penalty.

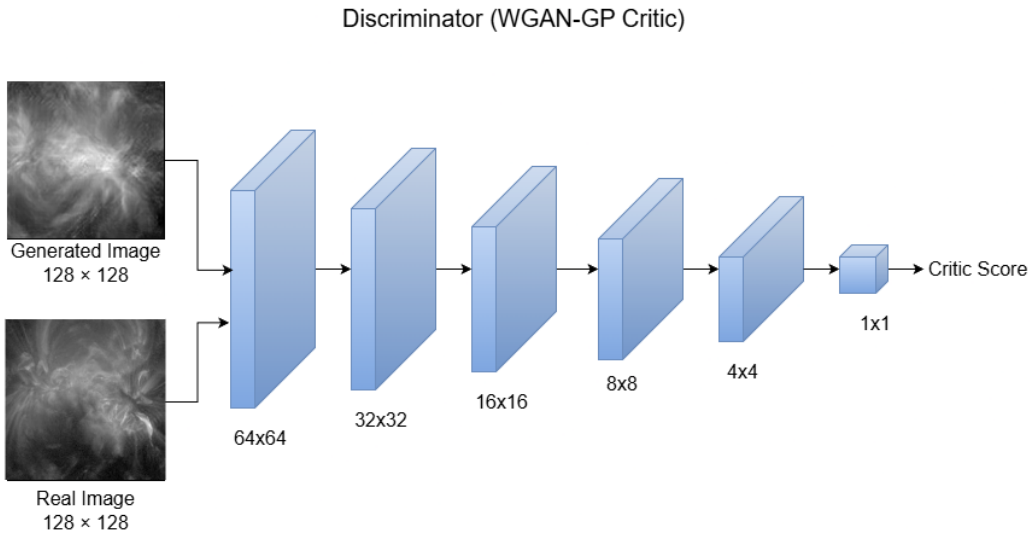

Figure 11: Training workflow of the WGAN-GP discriminator architecture.

Training follows the standard WGAN-GP objective. For a batch of real images x and generated images $x' = G(z)$, the critic loss is defined as

$$L_D = \mathbb{E}[D(x')] - \mathbb{E}[D(x)] + \lambda_{gp}\,\mathbb{E}\big[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2\big] \tag{4}$$

where $\hat{x}$ represents interpolated samples between real and generated images, and the gradient penalty coefficient λ_{gp} is set to 5.0. The generator is trained to minimize

$$L_G = -\mathbb{E}[D(G(z))], \tag{5}$$

which encourages the generation of samples that receive higher critic scores and are therefore closer to real severe flare images. Both the generator and the critic are optimized using the Adam optimizer with learning rates on the order of 10^{-4} and momentum parameters $(\beta_1, \beta_2) = (0.0, 0.9)$. To maintain stable training and accurate estimation of the Wasserstein distance, the critic is updated three times for each generator update ($n_{critic} = 3$). Training is performed on a GPU with automatic mixed precision enabled to improve computational efficiency. Model checkpoints are saved regularly over several hundred epochs until convergence is observed.

After training converges, the generator is used to produce synthetic severe-class images for each wavelength channel. Approximately $12,000$ synthetic

images are generated per wavelength, forming a pool of artificial severe flare samples. For each experiment, a fixed number of synthetic images is randomly drawn from this pool and combined with the real severe-class samples to produce different synthetic-to-real augmentation ratios (GAN percentages). Only the severe class is augmented, while the non-severe class distribution remains unchanged. The resulting augmented datasets are then used to train the corresponding single-wavelength classifier, enabling a systematic evaluation of the impact of GAN-based augmentation on severe flare prediction performance.

3.7 Evaluation Metrics

Model performance is evaluated using a set of classification metrics commonly adopted in solar flare classification, as summarized in Table 1. Particular emphasis is placed on metrics that are robust to severe class imbalance and suitable for rare-event classification, reflecting the operational importance of detecting high-impact solar flares. Among these metrics, the True Skill Statistic (TSS) is highlighted due to its insensitivity to class imbalance and its widespread use in space weather applications. Overall classification performance is further assessed using the ROC-AUC metric, which provides a threshold-independent measure of how well the model separates severe and non-severe events. Because several classification metrics depend explicitly on the choice of decision threshold, threshold selection is treated as a validation-based optimization step. Specifically, decision thresholds are selected by maximizing TSS on the validation set,

$$\mathrm{TSS}(\tau) = \mathrm{TPR}(\tau) - \mathrm{FPR}(\tau), \qquad \tau^\star = \arg\max_{\tau \in [0,1]} \mathrm{TSS}(\tau), \tag{6}$$

and are then fixed during evaluation on the test set to ensure fair and unbiased performance assessment.

In addition to classification performance, image-quality metrics are used to evaluate the realism of GAN-generated samples. Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) are employed to quantify pixel-level fidelity and structural similarity between real and synthetic images, respectively. These metrics provide complementary insight into the suitability of generated samples for data augmentation.

Table 1: Evaluation Metrics

Metric	Definition	Purpose
Classification Metrics		
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Overall correctness (sensitive to imbalance)
Precision	$\frac{TP}{TP+FP}$	Reliability of positive predictions
Recall (TPR)	$\frac{TP}{TP+FN}$	Detection of true flare events
F1 Score	$2\frac{PR}{P+R}$	Balance between precision and recall
TSS	$\frac{TP}{TP+FN} - \frac{FP}{FP+TN}$	**Skill independent of class imbalance**
ROC-AUC	Area under ROC curve	Threshold-independent discrimination
GAN Image Quality Metrics		
PSNR	$10 \log_{10} \left(\frac{MAX^2}{MSE} \right)$	Measures pixel-level fidelity between real and generated images
SSIM	Structural similarity index	Measures structural and perceptual similarity between images

4 Results

Before applying any data augmentation techniques, the baseline performance of the proposed fusion models is evaluated to establish a reference point for subsequent experiments. The residual neural fusion model achieves a TSS of 0.412 ± 0.024, as reported in Table 2. The recall is 0.819 ± 0.038, precision is 0.323 ± 0.022, and the ROC-AUC is 0.706 ± 0.012. The XGBoost-based fusion model achieves a higher baseline performance across most evaluation metrics, with a TSS of 0.500 ± 0.016, recall of 0.779 ± 0.055, precision of 0.399 ± 0.020, and a ROC-AUC of 0.750 ± 0.008. After establishing baseline performance, the quality of the images generated by the GAN is evaluated to assess their suitability for data augmentation. A qualitative comparison between real and generated AIA 131 images is presented in Fig. 12. Quantitative evaluation of the generated images is performed using PSNR and SSIM. The results for a selected wavelength are summarized in Table 3.

Table 2: Baseline performance of the proposed fusion models without GAN augmentation. Results are reported as mean ± standard deviation over three independent runs.

Metric	**Residual Fusion (ResNet)**	**XGBoost Fusion**
TSS	0.412 ± 0.024	0.500 ± 0.016
F1 Score	0.462 ± 0.017	0.526 ± 0.005
Recall	0.819 ± 0.038	0.779 ± 0.055
Precision	0.323 ± 0.022	0.399 ± 0.020
ROC-AUC	0.706 ± 0.012	0.750 ± 0.008
Accuracy	0.636 ± 0.043	0.732 ± 0.022

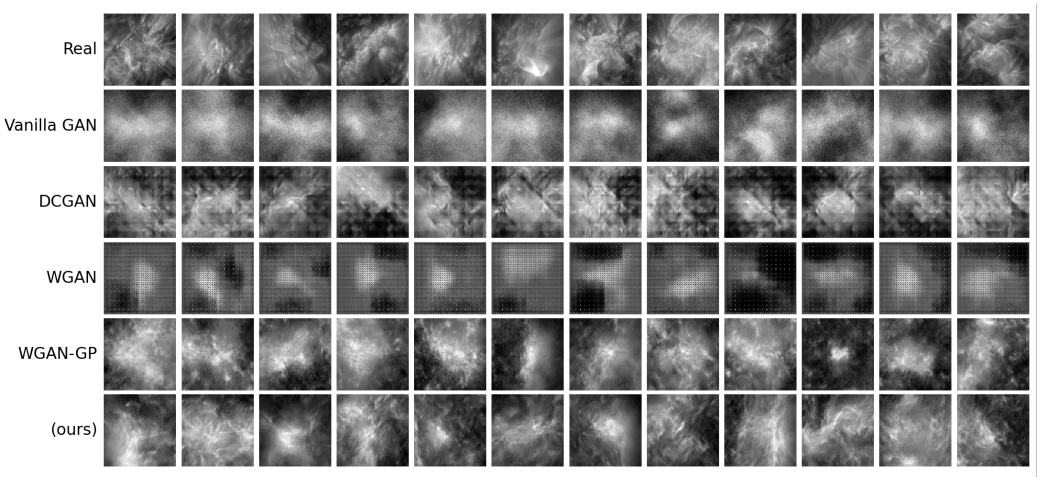

Figure 12: Qualitative comparison between real AIA 131 images and samples generated by different GAN variants.

Table 3: Quantitative comparison of different GAN variants using SSIM and PSNR on a single wavelength. Results are reported for both paired and nearest matching strategies.

GAN Model	Paired		Nearest	
	Mean SSIM	Mean PSNR	Mean SSIM	Mean PSNR
Vanilla GAN	0.306	16.797	0.391	19.293
DCGAN	0.247	15.178	0.331	17.831
WGAN	0.117	14.923	0.140	16.256
WGAN-GP	0.355	16.793	0.452	19.156
WGAN-GP + ResNet + Attention (ours)	**0.353**	**16.429**	**0.467**	**19.637**

The PSNR and SSIM values of the selected GAN across all wavelengths are reported in Table 4.

Table 4: PSNR and SSIM of the selected GAN across all wavelengths. Metrics are reported for paired and nearest matching strategies (N=128).

Wavelength	Paired		Nearest	
	Mean SSIM	Mean PSNR	Mean SSIM	Mean PSNR
AIA 94	0.394	16.174	0.526	19.781
AIA 171	0.291	15.591	0.440	17.956
AIA 193	0.337	14.331	0.526	18.710
AIA 131	0.353	16.279	0.505	20.448
AIA 211	0.438	14.969	0.597	19.274
AIA 304	0.436	18.601	0.571	21.447
AIA 335	0.530	15.395	0.672	21.738
AIA 1700	0.115	14.188	0.263	18.024
HMI continuum	**0.478**	**13.549**	**0.683**	**22.099**
HMI magnetogram	0.301	15.517	0.464	19.063

The effect of GAN-based data augmentation on the residual neural fusion model is evaluated using synthetic-to-real data ratios of 0%, 10%, 30%, 40%, and 50%. The corresponding performance metrics are reported in Table 5.

Table 5: Performance of the residual fusion model under different synthetic-to-real data ratios. Results are reported as mean ± standard deviation over three independent runs.

GAN Ratio	TSS	F1	Recall	Precision	ROC-AUC	Accuracy
0%	0.412 ± 0.024	0.462 ± 0.017	0.819 ± 0.038	0.323 ± 0.022	0.706 ± 0.012	0.636 ± 0.043
10%	0.435 ± 0.030	0.490 ± 0.013	0.725 ± 0.052	0.371 ± 0.010	0.718 ± 0.015	0.713 ± 0.015
30%	0.435 ± 0.017	0.492 ± 0.008	0.720 ± 0.086	0.379 ± 0.037	0.717 ± 0.008	0.715 ± 0.041
40%	0.436 ± 0.055	0.484 ± 0.029	0.764 ± 0.061	0.355 ± 0.026	0.718 ± 0.028	0.690 ± 0.027
50%	0.414 ± 0.035	0.484 ± 0.013	0.673 ± 0.067	0.380 ± 0.011	0.707 ± 0.018	0.728 ± 0.015

The performance of the XGBoost-based fusion model under different synthetic-to-real data ratios is summarized in Table 6.

Table 6: Performance of the XGBoost-based fusion model under different synthetic-to-real data ratios. Results are reported as mean ± standard deviation over three independent runs.

GAN Ratio	TSS	F1	Recall	Precision	ROC-AUC	Accuracy
0%	0.500 ± 0.016	0.526 ± 0.005	0.779 ± 0.055	0.399 ± 0.020	0.750 ± 0.008	0.732 ± 0.022
10%	0.495 ± 0.013	0.535 ± 0.010	0.749 ± 0.038	0.418 ± 0.013	0.751 ± 0.007	0.753 ± 0.010
30%	0.489 ± 0.020	0.519 ± 0.010	0.787 ± 0.042	0.383 ± 0.015	0.746 ± 0.011	0.719 ± 0.018

5 Discussion

The experimental results demonstrate that both fusion strategies are capable of learning meaningful representations for solar flare classification; however, they exhibit fundamentally different behaviors under data augmentation. At baseline, the residual neural fusion model prioritizes recall over precision, indicating a tendency to detect most flare events at the cost of increased false positives. Such behavior is often desirable in space weather applications, where missed flare events can have severe operational consequences. In contrast, the XGBoost-based fusion model achieves a more balanced trade-off between recall and precision, resulting in stronger overall discrimination performance without the use of synthetic data. This suggests that gradient-boosted decision trees are particularly effective at exploiting fixed feature embeddings, leading to stable generalization when trained exclusively on real observations.

The qualitative and quantitative evaluation of GAN-generated images reveals that the proposed WGAN-GP + ResNet + Attention architecture produces synthetic samples that closely preserve structural and textural characteristics of real solar observations. Although PSNR and SSIM values vary across wavelengths, consistently higher scores for both EUV and HMI channels indicate that the generator captures wavelength-specific features rather than overfitting to a single modality. This cross-channel consistency is not immediately apparent from visual inspection alone but is critical for multi-view fusion tasks.

When introducing GAN-based augmentation, the residual neural fusion model benefits from moderate levels of synthetic data. Performance improvements observed at intermediate augmentation ratios suggest that additional sample diversity enhances the model's ability to learn robust hierarchical representations. However, the degradation in performance at higher synthetic ratios indicates the presence of distributional bias, where excessive synthetic

data begins to dominate the training process and reduces generalization to real observations. Interestingly, the XGBoost-based fusion model does not exhibit similar gains from augmentation. The absence of performance improvement suggests that models relying on fixed embeddings are more sensitive to subtle shifts in feature distributions, even when synthetic samples are visually realistic. This highlights an important distinction between adaptive neural architectures and tree-based models in the context of data augmentation.

Despite the promising results, this work has several limitations. The experiments focus on binary flare classification and do not capture flare intensity levels or temporal evolution. Additionally, the dataset size remains limited. Future work will address these limitations by incorporating temporal modeling and multi-class flare prediction.

6 Conclusion

This paper presented *SunGuard*, a modular multi-view framework for severe ($\geq$M-class) solar flare classification using ten synchronized SDO channels from the SDOBenchmark dataset, where each wavelength is first modeled independently using EfficientNetV2-S to learn physically meaningful, wavelength-specific representations, and the resulting embeddings are then fused using two complementary feature-level strategies: a 1D residual neural fusion network and an XGBoost ensemble to capture cross-wavelength interactions in a controlled and computationally efficient manner; in addition, to address the severe class imbalance, we investigated GAN-based minority-class augmentation and selected WGAN-GP (with an attention-enhanced generator) due to its stability and its ability to preserve structural characteristics of solar observations, and we adopted validation-driven threshold optimization (maximizing TSS) to ensure fair and operationally relevant evaluation. Future work will extend this study beyond binary classification by incorporating temporal modeling of pre-flare evolution, exploring multi-class flare intensity prediction, scaling experiments to larger and more diverse datasets, and improving interpretability by localizing wavelength- and region-level cues (e.g., polarity inversion line structures) and integrating more physics-informed constraints into both the representation learning and generative augmentation components.

References

[1] "Solar Storms and Flares - NASA Science," Sep. 2024. [Online]. Available: https://science.nasa.gov/sun/solar-storms-and-flares/

[2] "Solar Flares (Radio Blackouts) | NOAA / NWS Space Weather Prediction Center." [Online]. Available: https://www.swpc.noaa.gov/phenomena/solar-flares-radio-blackouts

[3] "Remembering the Great Halloween Solar Storms," Nov. 2016. [Online]. Available: https://www.ncei.noaa.gov/news/great-halloween-solar-storm-2003

[4] O. Vural, S. M. Hamdi, and S. F. Boubrahimi, "Solar flare prediction using multivariate time series of photospheric magnetic field parameters: A comparative analysis of vector, time series, and graph data representations," *Remote Sensing*, vol. 17, no. 6, 2025.

[5] D. Boteler, "The super storms of august/september 1859 and their effects on the telegraph system," *Advances in Space Research*, vol. 38, no. 2, pp. 159–172, 2006, the Great Historical Geomagnetic Storm of 1859: A Modern Look.

[6] "BBC World Service - Witness History, What the 1989 solar storm did to Quebec." [Online]. Available: https://www.bbc.co.uk/programmes/w3ct4x8h

[7] "Near Miss: The Solar Superstorm of July 2012 - NASA Science," Jul. 2014. [Online]. Available: https://science.nasa.gov/science-research/planetary-science/23jul_superstorm/

[8] "Impact of Space Weather on Aviation." [Online]. Available: https://skybrary.aero/articles/impact-space-weather-aviation

[9] E. Schmölter, J. Berdermann, V. Wilken, and D. Wenzel, "Should we monitor space weather effects on surveillance technologies used in air traffic management?—first results," *Space Weather*, vol. 23, no. 4, p. e2025SW004352, 2025.

[10] "Safeguarding Satellites: How NOAA Monitors Space Weather to Prevent Disruptions," Dec. 2025. [Online]. Available: https:

//www.nesdis.noaa.gov/news/safeguarding-satellites-how-noaa-monitors-space-weather-prevent-disruptions

[11] C. J. Schrijver, "A characteristic magnetic field pattern associated with all major solar flares and its use in flare forecasting," *The Astrophysical Journal*, vol. 655, no. 2, p. L117, jan 2007.

[12] I. Kontogiannis, "The characteristics of flare- and cme-productive solar active regions," *Advances in Space Research*, vol. 71, no. 4, pp. 2017–2037, 2023, recent progress in the physics of the Sun and heliosphere.

[13] H. Sun, W. Manchester IV, and Y. Chen, "Improved and interpretable solar flare predictions with spatial and topological features of the polarity inversion line masked magnetograms," *Space Weather*, vol. 19, no. 12, p. e2021SW002837, 2021, e2021SW002837 2021SW002837.

[14] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. PMLR, 09–15 Jun 2019, pp. 6105–6114. [Online]. Available: https://proceedings.mlr.press/v97/tan19a.html

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[16] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 785–794.

[17] M. S. Wheatland, "A statistical solar flare forecast method," *Space Weather*, vol. 3, no. 7, 2005.

[18] G. Barnes and K. D. Leka, "Evaluating the Performance of Solar Flare Forecasting Methods," *Astrophysical Journal Letters*, vol. 688, no. 2, p. L107, Dec. 2008.

[19] D. S. Bloomfield, P. A. Higgins, R. T. J. McAteer, and P. T. Gallagher, "Toward Reliable Benchmarking of Solar Flare Forecasting Methods," *Astrophysical Journal Letters*, vol. 747, no. 2, p. L41, Mar. 2012.

[20] M. G. Bobra and S. Couvidat, "Solar Flare Prediction Using SDO/HMI Vector Magnetic Field Data with a Machine-learning Algorithm," *The Astrophysical Journal*, vol. 798, no. 2, p. 135, Jan. 2015.

[21] N. Nishizuka, K. Sugiura, Y. Kubo, M. Den, S. Watari, and M. Ishii, "Solar Flare Prediction Model with Three Machine-learning Algorithms using Ultraviolet Brightening and Vector Magnetograms," *The Astrophysical Journal*, vol. 835, no. 2, p. 156, Feb. 2017.

[22] K. Florios, I. Kontogiannis, S.-H. Park, J. Guerra, F. Benvenuto, D. Bloomfield, and M. Georgoulis, "Forecasting solar flares using magnetogram-based predictors and machine learning," *Solar Physics*, vol. 293, 01 2018.

[23] X. Huang, H. Wang, L. Xu, J. Liu, R. Li, and X. Dai, "Deep Learning Based Solar Flare Forecasting Model. I. Results for Line-of-sight Magnetograms," *The Astrophysical Journal*, vol. 856, no. 1, p. 7, Mar. 2018.

[24] X. Li, Y. Zheng, X. Wang, and L. Wang, "Predicting solar flares using a novel deep convolutional neural network," *The Astrophysical Journal*, vol. 891, no. 1, p. 10, feb 2020.

[25] V. Deshmukh, N. Flyer, K. Sande, and T. Berger, "Decreasing false-alarm rates in cnn-based solar flare prediction using sdo/hmi data," *The Astrophysical Journal Supplement Series*, vol. 260, p. 9, 05 2022.

[26] M. Li, Y. Cui, B. Luo, X. Ao, S. Liu, J. Wang, S. Li, C. Du, X. Sun, and X. Wang, "Knowledge-informed deep neural networks for solar flare forecasting," *Space Weather*, vol. 20, no. 8, p. e2021SW002985, 2022, e2021SW002985 2021SW002985.

[27] N. Nishizuka, Y. Kubo, K. Sugiura, M. Den, and M. Ishii, "Operational solar flare prediction model using deep flare net," *Earth, Planets and Space*, vol. 73, no. 1, p. 64, Mar 2021.

[28] R. Tang, W. Liao, Z. Chen, X. Zeng, J.-s. Wang, B. Luo, Y. Chen, Y. Cui, M. Zhou, X. Deng, H. Li, K. Yuan, S. Hong, and Z. Wu, "Solar flare prediction based on the fusion of multiple deep-learning models," *The Astrophysical Journal Supplement Series*, vol. 257, no. 2, p. 50, dec 2021.

[29] K. Pelkum Donahue and F. Inceoglu, "Forecasting solar flares with a transformer network," *Frontiers in Astronomy and Space Sciences*, vol. Volume 10 - 2023, 2024.

[30] L. F. L. Grim and A. L. S. Gradvohl, "Solar Flare Forecasting Based on Magnetogram Sequences Learning with Multiscale Vision Transformers and Data Augmentation Techniques," *Solar Physics*, vol. 299, no. 3, p. 33, Mar. 2024.

[31] S. Roy, J. Schmude, R. Lal, V. Gaur, M. Freitag, J. Kuehnert, T. Kessel, D. Hegde, A. Munoz-Jaramillo, J. Jakubik, E. Vos, K. Mandal, A. Asanjan, J. Almeida, T. Singh, K. Yang, C. Pandey, J. Hong, and R. Ramachandran, "Surya: Foundation model for heliophysics," Aug. 2025.

[32] "SDO | Solar Dynamics Observatory." [Online]. Available: https://sdo.gsfc.nasa.gov/

[33] "AIA - Atmospheric Imaging Assembly." [Online]. Available: https://aia.lmsal.com/

[34] "Helioseismic and Magnetic Imager for SDO." [Online]. Available: http://hmi.stanford.edu/

[35] R. B. Aerni, Michael, "SDOBenchmark | SDOBenchmark - Solar flare prediction image dataset." [Online]. Available: http://i4ds.github.io/SDOBenchmark/

[36] Y. Wang, Z. Pan, J. Zheng, L. Qian, and L. Mingtao, "A hybrid ensemble method for pulsar candidate classification," *Astrophysics and Space Science*, vol. 364, 08 2019.