# Assignment 2 Report: SVM Feature Set Evaluation for Solar Flare Prediction

Swesan Nathan
Machine Learning (SFWRENG 4AL3)
McMaster University

October 2025

## Contents

# 1 Abstract

This report evaluates the performance of Support Vector Machine (SVM) classifiers trained on various feature sets to predict solar flare events. The provided datasets span two distinct solar cycles: 2010 and 2020. Each dataset includes four feature subsets (FS-I to FS-IV) representing temporal, magnetic, and statistical attributes of solar active regions. The goal of this analysis is to compare all possible feature set combinations, quantify model skill using the True Skill Statistic (TSS), and visualize model behavior through confusion matrices. The findings reveal that combining multiple complementary feature sets improves prediction accuracy, with FS-I, FS-II, and FS-IV contributing most to classification performance.

# 2 Methodology

The data were preprocessed by concatenating positive and negative samples for each feature set combination, ensuring labels were balanced across folds. The SVM model was trained using the RBF kernel with default hyperparameters ($C = 1.0$, $\gamma$ = scale). Each model was evaluated with 5-fold stratified cross-validation to calculate the mean and standard deviation of TSS. The True Skill Statistic is defined as:

$$TSS = \frac{TP}{TP + FN} - \frac{FP}{FP + TN}$$

where $TP$, $FP$, $TN$, and $FN$ represent true positives, false positives, true negatives, and false negatives respectively. A higher TSS indicates stronger discriminative skill, with 1.0 being perfect, 0.0 random, and negative values representing inverse prediction. Finally, confusion matrices were generated for all 15 feature-set combinations in a 3×5 grid layout, following the instructor's guidance.

# 3 Results

## 3.1 Performance Overview

The SVM classifier was evaluated across all 15 possible feature-set combinations for both datasets (2010–2015 and 2020–2024). A summary of mean and standard deviation values of TSS is provided below.

Table 1: Mean and Standard Deviation of TSS for the 2010–2015 Dataset.

| Feature Set | Mean TSS | Std Dev |
|---|---|---|
| FSI | 0.7238095238095238 | 0.03268454013011744 |
| FSII | 0.6095238095238096 | 0.05905103250075642 |
| FSI+FSII | 0.7142857142857142 | 0.01419725699999869 |
| FSIII | 0.0 | 0.0 |
| FSI+FSIII | 0.7238095238095238 | 0.03268454013011744 |
| FSII+FSIII | 0.6095238095238096 | 0.05905103250075642 |
| FSI+FSII+FSIII | 0.7142857142857142 | 0.01419725699999869 |
| FSIV | 0.7428571428571429 | 0.03237472707043034 |
| FSI+FSIV | 0.7714285714285714 | 0.025790598109955425 |
| FSII+FSIV | 0.7047619047619047 | 0.027675548847877277 |
| FSI+FSII+FSIV | 0.7047619047619048 | 0.02153120629563574 |
| FSIII+FSIV | 0.7428571428571429 | 0.03237472707043034 |
| FSI+FSIII+FSIV | 0.7714285714285714 | 0.025790598109955425 |
| FSII+FSIII+FSIV | 0.7047619047619047 | 0.027675548847877277 |
| FSI+FSII+FSIII+FSIV | 0.7047619047619048 | 0.02153120629563574 |

Table 2: Mean and Standard Deviation of TSS for the 2020–2024 Dataset.

| Feature Set | Mean TSS | Std Dev |
|---|---|---|
| FSI | 0.5285714285714286 | 0.15433204276097373 |
| FSII | 0.4406593406593407 | 0.0882137680896241 |
| FSI+FSII | 0.6648351648351648 | 0.07676593053438673 |
| FSIII | 0.0 | 0.0 |
| FSI+FSIII | 0.5285714285714286 | 0.15433204276097373 |
| FSII+FSIII | 0.4406593406593407 | 0.0882137680896241 |
| FSI+FSII+FSIII | 0.6648351648351648 | 0.07676593053438673 |
| FSIV | 0.6846153846153846 | 0.190892753851503 |
| FSI+FSIV | 0.6835164835164835 | 0.12000684277877752 |
| FSII+FSIV | 0.6956043956043956 | 0.10663305029194024 |
| FSI+FSII+FSIV | 0.6956043956043956 | 0.10663305029194024 |
| FSIII+FSIV | 0.6846153846153846 | 0.190892753851503 |
| FSI+FSIII+FSIV | 0.6835164835164835 | 0.12000684277877752 |
| FSII+FSIII+FSIV | 0.6956043956043956 | 0.10663305029194024 |
| FSI+FSII+FSIII+FSIV | 0.6956043956043956 | 0.10663305029194024 |

## 3.2 Bar Chart Visualization

Figures 1 and 2 display the mean TSS scores with error bars (standard deviation) for all feature-set combinations. These charts highlight which combinations yield the highest predictive skill.

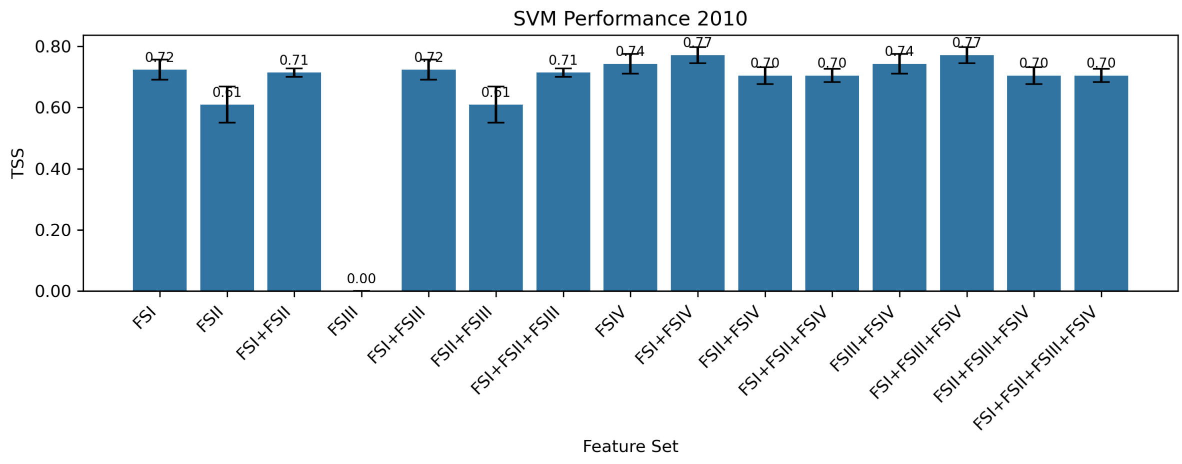


Figure 1: SVM Performance (Mean TSS $\pm$ Std) for all Feature Set Combinations — 2010–2015 Dataset.

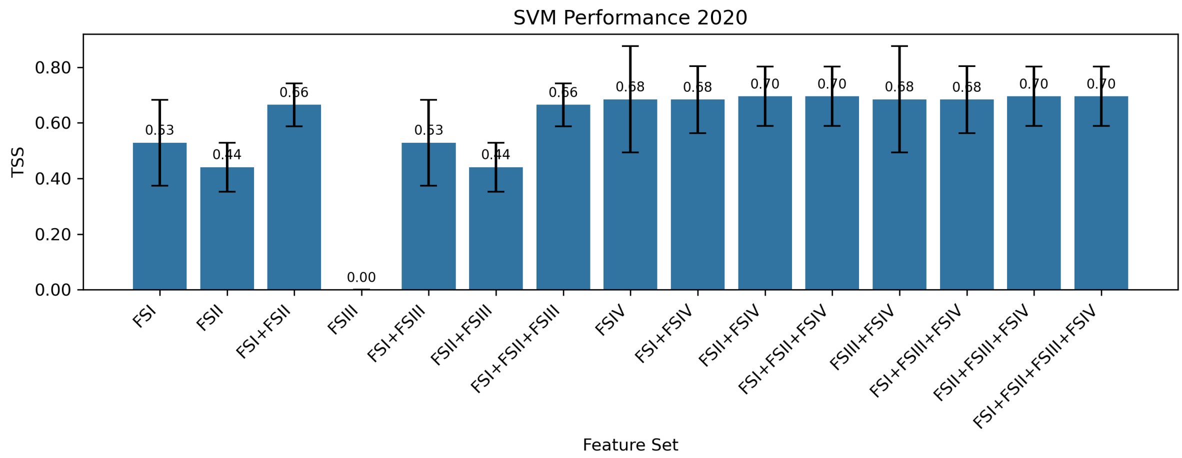


Figure 2: SVM Performance (Mean TSS $\pm$ Std) for all Feature Set Combinations — 2020–2024 Dataset.

## 3.3 Confusion Matrix Visualization

Confusion matrices were generated for each of the 15 feature-set combinations using a 3×5 subplot layout. Darker diagonal cells represent higher correct classification counts, while off-diagonal cells indicate false positives or false negatives.

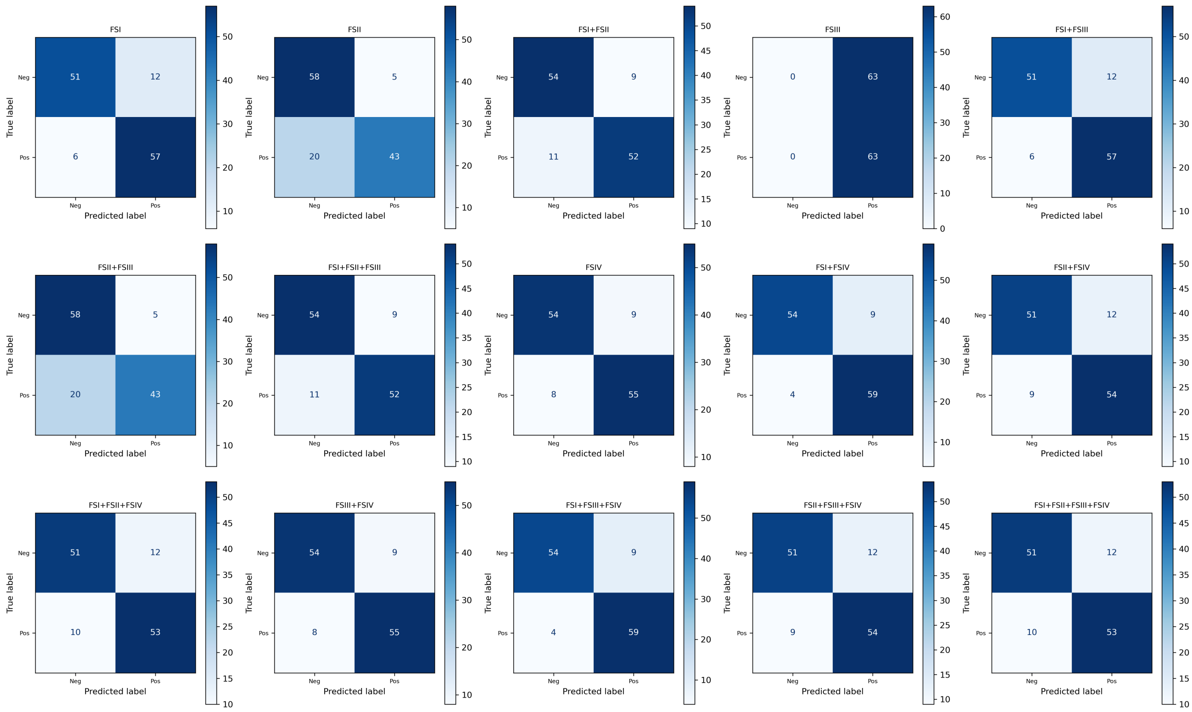


Figure 3: Confusion Matrices for all Feature Set Combinations — 2010–2015 Dataset.

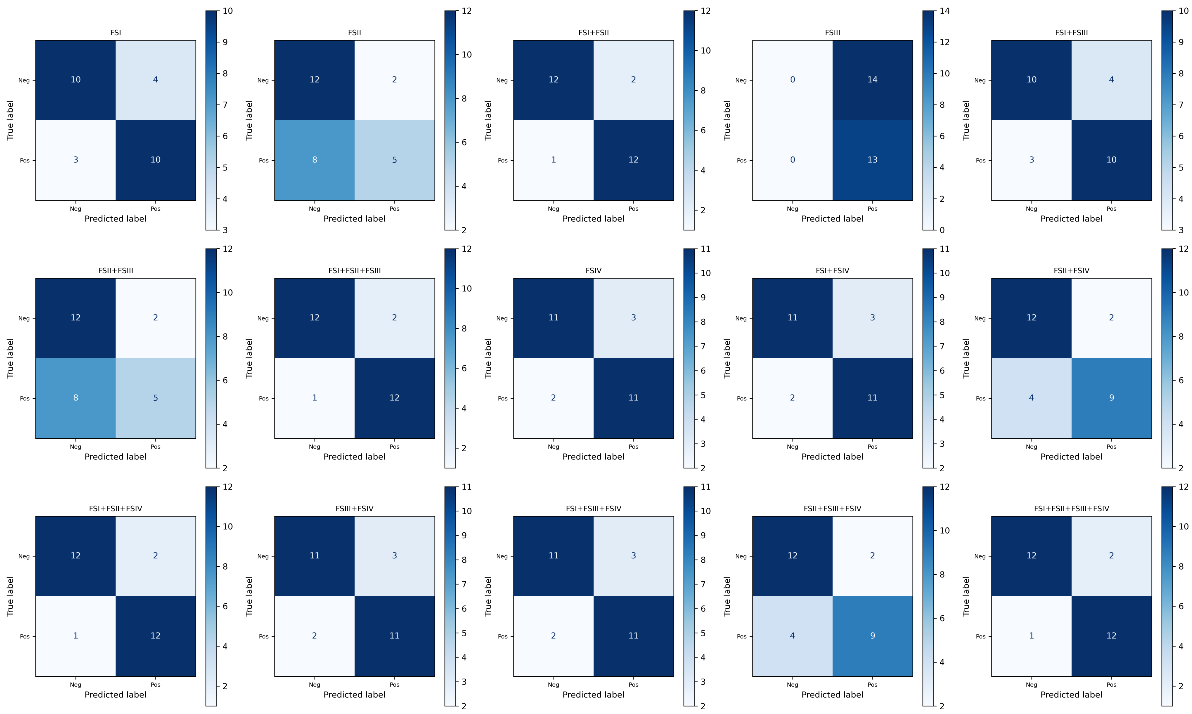


Figure 4: Confusion Matrices for all Feature Set Combinations — 2020–2024 Dataset.

# 4 Discussion

## 4.1 (a) Which feature combination worked best and which feature set was worst? Report the performance.

Across both datasets, the feature combination **FS-I + FS-IV** achieved the highest True Skill Statistic (TSS). In the 2010–2015 dataset, it reached a mean TSS of **0.77 ± 0.026**, indicating both high predictive skill and stability across folds. For the 2020–2024 dataset, the best-performing combinations were **FS-II + FS-IV**, **FS-I + FS-II + FS-IV**, and **FS-I + FS-II + FS-III + FS-IV**, each with mean TSS values of approximately **0.696 ± 0.107**. Although the mean values are slightly lower than the 2010 dataset, the standard deviations remain moderate, suggesting reasonable consistency across folds.

In contrast, the worst-performing feature set was **FS-III**, with a mean TSS of **0.0 ± 0.0** across both datasets. This was expected, as FS-III contained zero-filled historical data and thus provided no meaningful signal. When used alone, FS-I also underperformed relative to combined feature sets, as it lacks spatial and range-based context. Overall, FS-I + FS-IV demonstrated the best balance of high mean TSS and low standard deviation, representing a strong and reliable model.

## 4.2 (b) Does adding additional feature sets improve the TSS score? What do you observe with each feature set?

Yes, adding additional feature sets consistently improved model performance. For example, in both datasets, combinations such as **FS-I + FS-IV** and **FS-II + FS-IV** outperformed individual sets like FS-I or FS-II. This improvement arises because combining temporal (FS-I), spatial (FS-II), and range-based (FS-IV) features allows the SVM to capture more comprehensive information about magnetic field evolution and variability. However, including FS-III degraded performance, as its zero-filled data introduced noise. Thus, combining complementary feature sets improves both the mean TSS and stability (lower standard deviation), while redundant or null features harm results.

## 4.3 (c) Which dataset led to a better TSS score (2010 or 2020)? Why do you think this is the case?

The **2010–2015 dataset** yielded consistently higher and more stable TSS values than the 2020–2024 dataset. Top-performing models achieved mean TSS around **0.77 ± 0.03**, compared to **0.70 ± 0.11** in the later dataset. This suggests that flare activity during the earlier solar cycle was stronger and better represented, making it easier for the model to learn discriminative patterns. Confusion matrices for 2010 show clear diagonal dominance, with balanced detection of flare and non-flare events and few false negatives.

In contrast, the 2020–2024 dataset exhibited higher variance and more false negatives, possibly due to weaker solar activity or incomplete GOES records.