



ORIGINAL ARTICLE

Feature Reduction on Sym-H Index Image Using Principal Component Analysis Approach

Aznilinda Zainuddin¹, Muhammad Asraf Hairuddin^{1*}, Zatul Iffah Abd Latiff¹, Nur Dalila Khirul Ashar¹, Anwar Santoso², Mohamad Huzaimy Jusoh³, Ahmad Ihsan Mohd Yassin⁴

¹Electrical Engineering Studies, College of Engineering, Universiti Teknologi MARA, Johor Branch, Pasir Gudang Campus, Masai, Malaysia.

²Center for Space, National Research and Innovation Agency, Bandung, Indonesia.

³School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia.

⁴Microwave Research Institute, Universiti Teknologi MARA, Shah Alam, Malaysia.

*Corresponding author: masraf@uitm.edu.my

Received: 11/12/2024, Accepted: 16/04/2025, Available Online: 30/04/2025

Abstract

Geomagnetic storms pose significant risks to technological systems on Earth. One of the ways to identify the level of a storm is from the Sym-H plot images. The fewer features used for image interpretation, the simpler and more efficient the analysis becomes. In this study, we applied Principal Component Analysis (PCA) to the Sym-H index images, initially consisting of seven statistical features. Through PCA, this study managed to reduce these features to just two principal components, capturing over 98% of the total variance in the first two components, thereby retaining essential information while simplifying the dataset. This reduction not only simplifies the visualization and interpretation of the Sym-H plot images but also retains the critical information necessary for understanding geomagnetic storm dynamics. By focusing on these two principal components, we can effectively present and analyse the essential patterns and behaviours of geomagnetic activity during storm events. The findings highlight the potential of PCA to enhance space weather forecasting and improve the resilience of technological infrastructure against solar storm impacts.

Keywords: Principal Component Analysis (PCA), Sym-H Index, Geomagnetic Storms, Space Weather Forecasting, Feature Reduction

Introduction

Geomagnetic storms can be deemed one of the destructive forms of Space Weather event, adversely affecting various technological systems on the Earth. Geomagnetic storms cause fluctuations in the Earth's magnetic field, and the intensity of these variations is represented by the Sym-H index (Collado-Villaverde et al., 2021). Sym-H index is also used in examining how the ionosphere responds to the solar events with a view of determining the effect of space weather

on the earth's magnetic field and ionosphere (Younas et al., 2018). Sym-H index is employed as the operative Dst index with higher time resolution for assessing the geomagnetic storm intensification (Wabliss et al., 2006). However, the complexity and high dimensionality of the data associated with Sym-H images can pose significant challenges for analysis and interpretation.

Principal Component Analysis (PCA) is one of the popular techniques applied in computer science, mathematics, and artificial intelligence for reducing the dimensionality of the data (Karamizadeh et al., 2013). It entails the process of converting the original data set that has many variables, complexity and high dimensionality into a new set of variables which are orthogonal and yet retains as much of the variance of the original data as possible. In other words, PCA reduces the dimensionality of data by projecting them into a lower-dimensional linear subspace, and at the same time, it preserves the important information in the data (Pedretti et al., 2023). This reduction in dimensionality is important when working with large multivariate data sets since it assists in the simplification of the data structure and analysis (Jolliffe, 2014). However, despite its popularity, PCA has several drawbacks that should be noted. Wei (Wei and Ouyang, 2024) points out that classic dimensionality reduction approaches, such as PCA, entirely neglect the target information. This means that PCA focusses primarily on the variance of the data and may not account for the target variable's special qualities or links to the data. Another limitation of PCA is that it may be ineffective in finding weak features in time-varying signals or data with low correlations (Shi et al., 2020). The approach of extracting principal components in PCA includes picking only a small number of primary elements, which might result in information loss and make it difficult to find subtle or weak patterns in the data.

In the field of machine learning and data analysis, PCA is particularly valuable in scenarios where a dataset contains many variables that are highly correlated with each other. This correlation means that the variables share similar information, leading to redundancy and complexity in the data. PCA addresses this issue by transforming the correlated variables into a smaller set of uncorrelated variables known as principal components. PCA is widely used to increase the performance of the algorithms by eliminating unimportant features or variables while preserving the features that contain important information (Reddy et al., 2020). PCA has also been reported to enhance the deep neural network models when used in tasks such as house price predictor (Mostofi et al., 2021). Furthermore, PCA has been compared with other features extraction methods such as Linear Discriminant Analysis (LDA) for other machine learning algorithms and the results have been compared (Reddy et al., 2020). Some of the areas which have utilized PCA include agriculture, climate science, and medical research among others, which has demonstrated the capability of PCA in working with various kinds of data (Ali et al., 2023, Rieger & Levang, 2024). This demonstrates how the PCA and related methods are versatile in handling and analysing of complex data and modelling them.

PCA is commonly applied in space weather studies for the analysis and interpretation of multivariate data concerning different aspects of space weather. For example, Benitez et al. Benitez (Benitez et al., 2024) used PCA to reduce the feature correlation of weather variables to effectively analyse the solar photovoltaic production efficiency. Likewise, Wang et al., (2022) revealed the efficacy of PCA in analysing the quantitative variables that are inter-correlated for meteorological visibility prediction. Furthermore, PCA has been used to predict TEC changes in certain areas; thus, it can be used to predict ionosphere changes during geomagnetic storms and solar flares (Morozova et al., 2022). The PCA based model shows that most of the TEC changes can be explained by the first two principal components. In the other work, Morozova and Rebbah (2023) investigates the application of PCA to identify the solar quiet (Sq) variation from the geomagnetic field records. The authors apply PCA to the X, Y, and Z components of the geomagnetic field and different time intervals and solar and geomagnetic activity levels. It is

observed that PCA is quite useful for extracting Sq variations especially for the Y and Z components of the geomagnetic field and the first principal component (PC1) is seen to represent Sq variations most of the time. High-Intensity Long-Duration Continuous AE Activity (HILDCAAs) was also modelled during the solar minimum of cycle 23/24 using PCA (Klausner et al., 2021). PCA is used on the geomagnetic indices to obtain the PCs that describe the main features of the geomagnetic field data variation and reveals that PCA is useful in distinguishing the dominant components of geomagnetic activity associated with HILDCAAs.

This study will leverage Principal Component Analysis (PCA) for feature reduction on Sym-H index plot images, which is used to interpret geomagnetic storm activity. By reducing the initial statistical features, the reduced dimensionality will help in maintaining essential information while providing a more manageable and optimum dataset for interpreting severe and weak geomagnetic storms via the Sym-H images.

Materials and Methods

Data Selection

Sym-H can be classified into three categories, weak storms SYM-H between -150 and -80 nT, moderate storms Sym-H between -300 and -150 nT and intense storms SYM-H of less than -300 nT (Hutchinson et al., 2011). However, in this work, the data are focused on severe and weak geomagnetic storms, with severe storms having the values of Sym-H between -100 nT to -600 nT. While for weak geomagnetic storms, the minimum Sym-H values are between -20 to -70 nT as shown in **Error! Reference source not found.** Five samples of Sym-H images show severe storms, and six Sym-H images represent weak storms. Data are retrieved from the OMNI database at https://omniweb.gsfc.nasa.gov/form/omni_min.html.

Table 1. Geomagnetic storms event selected with severe and weak level of storm.

Storm no.	Start date	End date	Minimum Sym-H (nT)	Category ^a
1	16 March, 2013	20 March, 2013	-132	S
2	16 March, 2015	20 March, 2015	-234	S
3	21 June, 2015	24 June, 2015	-208	S
4	6 September, 2017	10 September, 2017	-146	S
5	10 May, 2024	13 May, 2024	-518	S
6	1 July, 2019	4 July, 2019	-28	W
7	3 October, 2019	6 October, 2019	-22	W
8	10 December, 2021	13 December, 2021	-21	W
9	10 May, 2020	13 May, 2020	-22	W
10	20 June, 2021	23 June, 2021	-22	W
11	4 May, 2024	7 May, 2024	-66	W

^a S – Severe, W – Weak

Methodology

To achieve the objective of this study, the method is described in the flow chart provided in **Figure 1**, which entails the gathering and selecting of data, focusing on geomagnetic storm events of severe and weak intensity. Data is gathered and then temporal variations and characteristics related to geomagnetic storms is visualized by generating the time series plots. The essential statistical features from these images are then extracted and differentiated between the severe and weak Sym-H images. The features selected for PCA include mean, variance, skewness, kurtosis, variance intensity, number of peaks, and number of troughs. These features were chosen because they represent the central tendency, variability, distribution shape, and dynamic characteristics of the Sym-H index, which are essential in distinguishing between different intensities of geomagnetic storms.

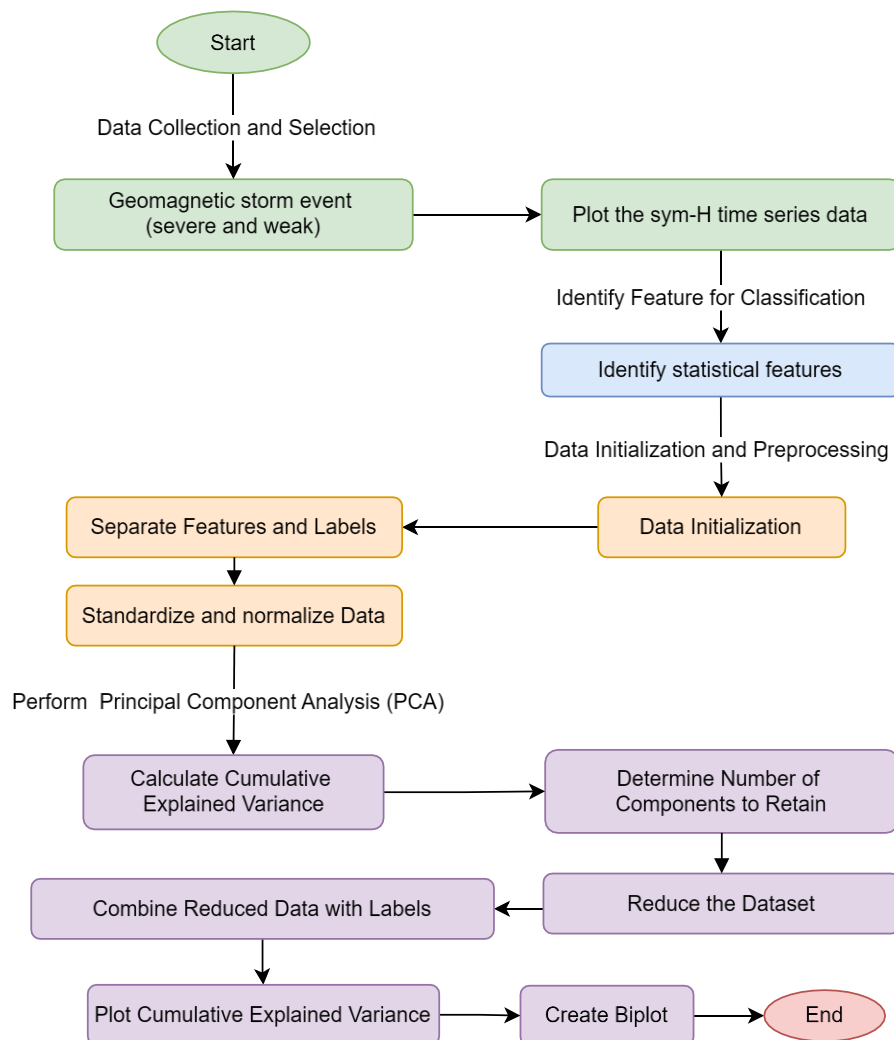


Figure 1. Research Method Flow for Feature Reduction on Sym-H Index Images Using PCA.

Organizing the data constitutes the steps of data initialization and data preprocessing. This step involves some data pre-processing techniques, which comprises of feature selection, extracting the input variables and labelling the output categories. There is a need to scale and normalize the data to ensure that all the features are utilizing the same measure in PCA. This normalization also helps to avoid the domination of any features by their scales since this can disproportionately influence the analysis. Following data standardization, PCA is carried out on the standardized data so that the dataset's dimensionality can be reduced while at the same time preserving as much variance as possible.

With the dataset being appropriately normalized, the variance in the data is significant enough to allow PCA to effectively reduce the dimensionality while preserving the most critical information. The PCA then performed the correlation matrix since the features vary in scale, to ensure that all features contribute equally to the analysis. This reduction is done by transforming the original features into a new data set that has two distinct variables known as the principal components. Subsequent to this, the total percent of variance is computed in order to identify how many of the principal components should be used to explain as much variation as possible from the data. In order to identify the number of components to retain, the accumulated explained variance is plotted. Thus, from the explained variance, the number of principal components to be retained is deduced and the rest of the dataset is then truncated out. The final step involves the construction of the biplots to facilitate the interpretation of the PCA outcomes. Biplot visualization helps unravel the structure of the data and the feat achieved by the PCA in the differentiation of various kinds of geomagnetic storms.

Results and Discussions

Statistical Features in Sym-H Index Time Series Plot Images

Figure 2 shows the image sample of the Sym-H Index time series plot image with all images in 3216 x 2461 pixels. From these images, various statistical features are extracted and tabulated in **Table 2** for weak and severe geomagnetic storms. These features include mean, variance, skewness, kurtosis, variance intensity, number of peaks, and number of troughs. The insights into the characteristics of the geomagnetic storms were gained by examining these features.

The mean values for both weak and severe storms are quite similar, hovering around 0.98. Even though the central tendency of the Sym-H index does not differ significantly, we identify all images that have a mean of 0.98 and above as weak storms. The variance on the other hand is slightly higher for severe storms (e.g., 0.020984) compared to weak storms (e.g., 0.017391). This suggests that severe storms exhibit greater variability in their Sym-H index values with 0.019 as its threshold. While for the skewness and kurtosis, it provides information about the distribution shape of the Sym-H index data. Kurtosis is a measure of the tailedness or the extremity of outliers in a probability distribution. It is calculated as follows:

$$\text{Kurtosis} = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Where:

n is the number of data points

x_i is each individual data point

$\bar{x}$ is the mean of the data points

s is the standard deviation of the data points

This formula provides the "excess kurtosis," which makes the kurtosis of a normal distribution zero. Positive values indicate a distribution with heavier tails and a sharper peak, while negative values indicate a distribution with lighter tails and a flatter peak. Both weak and severe storms show negative skewness, indicating a left-skewed distribution. The skewness values for severe storms (e.g., -6.46881) are generally less negative than for weak storms (e.g., -7.31453), implying a less pronounced skew in severe storms. Kurtosis values are higher for weak storms (e.g., 54.50229) than for severe storms (e.g., 42.84556), indicating that weak storms have heavier tails, and a higher peak compared to severe storms.

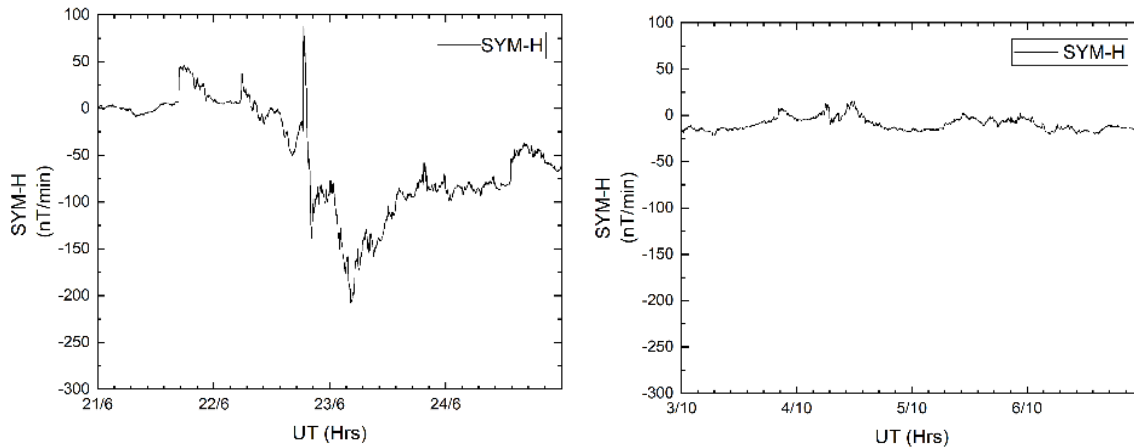


Figure 2. Samples of Sym-H time series plot of severe (on the left) and weak (on the right) geomagnetic storm.

Table 2. The statistical features of the Sym-H index plot

	Mean	Variance	Skewness	Kurtosis	Variance Intensity	Num Peaks	Num Troughs
Weak	0.982296	0.017391	-7.31453	54.50229	1588380	3	4
	0.983689	0.016045	-7.63719	59.32663	1533404	3	4
	0.982055	0.017623	-7.26249	53.74374	1573959	3	4
	0.982418	0.017273	-7.34127	54.89428	1555761	3	4
	0.983133	0.016582	-7.5037	57.30548	1240896	3	4
	0.982097	0.017582	-7.27152	53.87505	1583538	3	4
	0.982097	0.017582	-7.27152	53.87505	1583538	3	4
Severe	0.978556	0.020984	-6.6072	44.65505	1975188	9	10
	0.97769	0.021812	-6.46881	42.84556	1750736	5	6
	0.979737	0.019853	-6.80959	47.3705	1655304	5	6
	0.976816	0.022647	-6.33689	41.15611	2290367	15	16
	0.979414	0.020162	-6.75257	46.59726	1768275	7	8

Furthermore, the variance intensity, that is the intensity of the variance over the time series can be a specialized metric, often involving the change in variance over segments of the time series. This feature helps in detecting periods of high variability, which might be significant for the

phenomenon being studied, representing the overall intensity of variability. It is significantly higher in severe storms (e.g., 1975188) compared to weak storms (e.g., 1588380). This aligns with the expectation that severe storms should exhibit greater overall intensity and variability. Severe storms have more fluctuations in the levels, such as peaks (e.g. 9) and troughs (e.g. 10) compared to weak storms which have few fluctuations such as 3 peaks and 4 troughs. This implies that severe storms are characterized by more fluctuation and dynamism in the value of the Sym-H index.

Analysing the data containing the value of the Sym-H index, the main statistical characteristics significantly differentiate between weak and extreme storms. In addition, weak storms contain a lesser variance, have negative skewness, and have higher kurtosis as compared to severe storms, these storms have intense variance and possess a greater number of peaks and troughs than the weak storms. Collectively, these features offer a clear and distinct basis for identifying and differentiating between the characteristics and intensities of weak and severe geomagnetic storms, thereby facilitating accurate categorization.

Principal Component Analysis (PCA)

The biplot and the cumulative explained variance plot of the PCA results depicted a good indication of the higher-order structure of the analysed data in classifying Sym-H indices into intense or weak storms, and this is presented in **Figure 3**. The obtained biplot shows the projection of the variables based on the first two PCs (PC1 and PC2). Analysing the factor loadings, it is observed that for Component 2, the variables 'num Troughs,' 'num Peaks,' as well as the 'variance Intensity,' are significant and all have positive values, proportionally relating to the selected factor. Thus, it may be concluded that, although the 'Variance' and 'skewness' have negative values with Component 2, it means they have positive magnitude of linkage to Component 1, which indicates that they are more relevant to the variance reflected by this component. Next, the factor 'Kurtosis' and 'mean' have a high positive number on Component 1 meaning that they are strongly related to this component.

The second important remark is the interconnections of the variables involved. These are observed where such an assumption is made as with 'num Troughs' and 'num Peaks' where the latter is assumed to be a continuation of the former as it deals with a similar aspect of the data input. 'Variance' and 'skewness' exhibit a strong negative correlation with 'kurtosis' and 'mean,' indicating an inverse relationship between these sets of variables. In terms of the data distribution, the distribution of the red points (representing individual observations) indicates how well the principal components separate the data. The spread along the axes suggests variability within the dataset and highlights the significance of each component in capturing this variability.

The cumulative explained variance plot shows the percentage of variance explained by the principal components. In terms of the explained variance, the first two components explain over 98% of the total variance in the data, with Component 1 explaining around 91% and Component 2 adding approximately 7%. The inclusion of additional components beyond the second one contributes negligibly to the total explained variance, indicating that the first two components are sufficient to capture most of the variability in the dataset. While on dimensionality reduction, the steep rise in the explained variance with the first two components justifies the focus on these components for analysis, as they encapsulate the essential information needed for classifying the Sym-H indices. The percentage of variance captured by PCA shows that a major proportion of the variability in the dataset is accounted for by the first two PC's which points out to be useful input in categorizing different storms. The high variance indicated by the above

components is quite high, which confirms the effectiveness of PCA when applied to this dataset as it simplifies the dimensionality of the images.

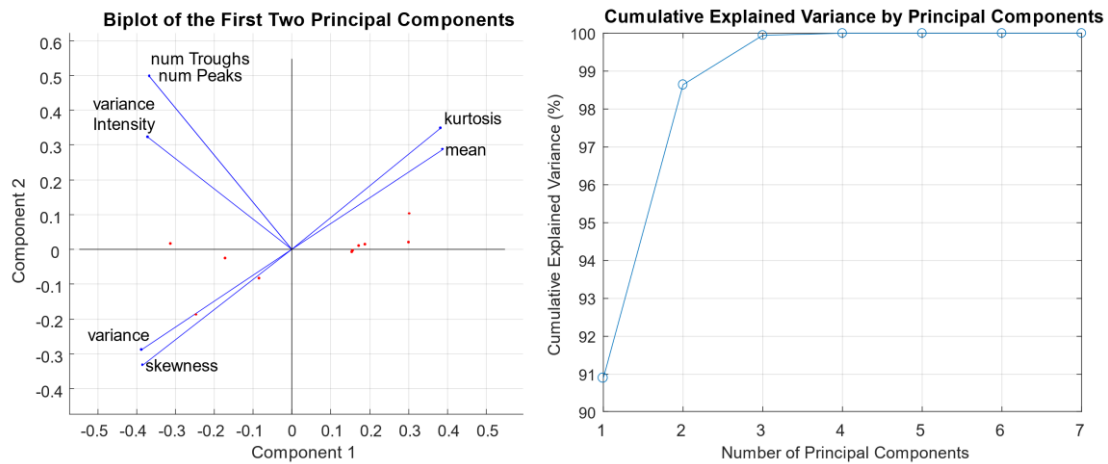


Figure 3. Principal Component Analysis (PCA) results of Sym-H index features. The biplot (left) shows the projection of the variables onto the first two principal components, highlighting their contributions and relationships. The cumulative explained variance plot (right) indicates that the first two components capture over 98% of the total variance, demonstrating the effectiveness of PCA in reducing dimensionality while retaining essential information for classifying solar storm categories.

Parameters like 'num Troughs,' 'num Peaks,' 'Intensity,' 'Variance,' 'Skewness,' 'Kurtosis,' and 'Mean' are central to determining the principal components, which therefore shows why they are relevant in characterizing the Sym-H index. The relations of these variables reveal the necessary patterns and dependencies that are vital in identifying the image characteristic. In other words, when carrying out the analysis, the researchers should concentrate on the first two principal components that will help in simplifying the data while keeping essential details, thus sparing resources and increasing the likelihood of correct categorization of the storm effects. Thus, it contributes more to interpretability and robustness of this analysis, and it can be used as a tool for understanding the space weather events' classification.

Error! Reference source not found. shows the new data sets that were generated after applying PCA on the Sym- H index data set having retained the first two principal components. This table shows how the given high dimensionality data is converted to a lower dimensionality without much loss of information in regard to the intensities of solar storms. Every row in **Error! Reference source not found.** is an observation, identified as either weak or severe geomagnetic storms. The values in the columns labelled 'Weak' and 'Severe' are the projections of the data points in the new dimension formed by the two principal components. These values show how each observation is mapped into the principal components and give a clear separation of one storm intensity from another. This projection demonstrates the feasibility of applying PCA to model the essential variations in the range of data sets.

The value of the principal components derives from their capability in expressing most of the original data's variation in terms of some linearly independent variables. The values listed in the **Error! Reference source not found.** represent the same set of points in the lower dimensional space and it is observed that weak and severe storm data points are separated more clearly. This clear separation indicates that PCA serves the purpose of improving from classifying

the different storm intensity with adequate distinction. Moreover, the decrease in the number of features preserves the distribution and data necessary for Sym-H category to the best extent,

Table 3. The new datasets with the number of components to retain is two.

Category	PC1	PC2
Weak	1.524606	0.091768
	2.673226	0.907071
	1.368377	-0.071665
	1.661599	0.125605
	2.653571	0.181046
	1.385487	-0.034421
	1.385487	-0.034421
Severe	-2.757631	0.138673
	-2.168982	-1.659497
	-0.746252	-0.734471
	-5.462024	1.310918
	-1.517463	-0.220604

*PC = Principal Component

despite the dimensionality of the problem. This does not only make the requirement of computational resources lesser but also increases the rate of analysis and classifications.

Conclusion

This study has effectively applied Principal Component Analysis (PCA) on the Sym-H index plot images that contained seven statistical features and was able to compress the features to two principal components. Both these factors accounted for more than 98% of the total variance, which means that the important information was kept, while the number of variables was decreased. Due to reduced dimensions, the main aspects of Sym-H plot images could be depicted and analysed in a better approach without losing the significant features that define the geomagnetic storm. This reduction demonstrates the efficiency of PCA in dealing with high dimensional space weather data to ease the analysis and increase the interpretability of the indices.

By focusing on the principal components, this study provides a clearer and more efficient method to analyse geomagnetic activity during storm events. Thus, the presented study highlights the possibilities of PCA application for the advancement of space weather classification and protection of technological systems against solar storms. Further work will include the establishment of these PCA reduced features into classification models to improve the accuracy of storm classification and forecasting.

Acknowledgments

The authors would like to thank Universiti Teknologi MARA (UiTM) and the Malaysian Ministry of Higher Education (MoHE) for the support given. The results presented in this paper rely on data collected at magnetic observatories. The authors thank the national institutes that support them and INTERMAGNET for promoting high standards of magnetic observatory practice. The authors acknowledge the NASA/GSFC Space Physics Data Facility's OMNIWeb service for geophysical

parameters data. This work is funded by Universiti Teknologi MARA (UiTM) under Grant No. 600-RMC 5/3/GPM (047/2023).

References

- A. Collado-Villaverde, P. Muñoz, and C. Cid, "Deep Neural Networks With Convolutional and LSTM Layers for SYM-H and ASY-H Forecasting," *Sp. Weather*, vol. 19, no. 6, Jun. 2021, doi: 10.1029/2021SW002748.
- A. Morozova and R. Rebbah, "Principal component analysis as a tool to extract Sq variation from the geomagnetic field observations: Conditions of applicability," *MethodsX*, vol. 10, no. January, p. 101999, 2023, doi: 10.1016/j.mex.2023.101999.
- A. Morozova, T. Barata, T. Barlyaeva, and R. Gafeira, "Total Electron Content PCA-NN Prediction Model for South-European Middle Latitudes," *Atmosphere (Basel)*, vol. 14, no. 7, p. 1058, Jun. 2023, doi: 10.3390/atmos14071058.
- B. I. Jolliffe, "Principal Component Analysis," pp. 1–5, 2014.
- B. Pedretti *et al.*, "Experimental Assessment of PCA and DT Classification for Streamlined Position Reconstruction in Anger Cameras," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 7, no. 4, pp. 315–325, 2023, doi: 10.1109/TRPMS.2022.3212616.
- F. Mostofi, V. Toğan, and H. B. Başağa, "House price prediction: A data-centric aspect approach on performance of combined principal component analysis with deep neural network model," *J. Constr. Eng. Manag. Innov.*, vol. 4, no. 2, pp. 106–116, Jun. 2021, doi: 10.31462/jcemi.2021.02106116.
- G. T. Reddy *et al.*, "Analysis of Dimensionality Reduction Techniques on Big Data," *IEEE Access*, vol. 8, pp. 54776–54788, 2020, doi: 10.1109/ACCESS.2020.2980942.
- H. Shi, J. Guo, X. Bai, L. Guo, Z. Liu, and J. Sun, "Gearbox Incipient Fault Detection Based on Deep Recursive Dynamic Principal Component Analysis," *IEEE Access*, vol. 8, pp. 57646–57660, 2020, doi: 10.1109/ACCESS.2020.2982213.
- I. Ali, Z. Mushtaq, S. Arif, A. D. Algarni, N. F. Soliman, and W. El-Shafai, "Hyperspectral Images-Based Crop Classification Scheme for Agricultural Remote Sensing," *Comput. Syst. Sci. Eng.*, vol. 46, no. 1, pp. 303–319, 2023, doi: 10.32604/csse.2023.034374.
- I. B. Benitez, K. I. Repedro, and J. A. Principe, "ASSESSMENT OF SOLAR PV OUTPUT PERFORMANCE WITH VARYING TILT ANGLES AND WEATHER DATA FROM ERA5: CASE OF MUNTINLUPA CITY, PHILIPPINES," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLVIII-4/W, no. 4/W8-2023, pp. 47–52, Apr. 2024, doi: 10.5194/isprs-archives-XLVIII-4-W8-2023-47-2024.
- J. A. Hutchinson, D. M. Wright, and S. E. Milan, "Geomagnetic storms over the last solar cycle: A superposed epoch analysis," *J. Geophys. Res. Sp. Phys.*, vol. 116, no. A9, p. n/a-n/a, Sep. 2011, doi: 10.1029/2011JA016463.
- J. A. Wanliss and K. M. Showalter, "High-resolution global storm index: Dst versus SYM-H," *J. Geophys. Res. Sp. Phys.*, vol. 111, no. A2, Feb. 2006, doi: 10.1029/2005JA011034.
- N. Rieger and S. J. Levang, "xeofs: Comprehensive EOF analysis in Python with xarray," *J. Open Source Softw.*, vol. 9, no. 93, p. 6060, 2024, doi: 10.21105/joss.06060.
- S. Karamizadeh, S. M. Abdullah, A. A. Manaf, M. Zamani, and A. Hooman, "An Overview of Principal

Component Analysis,” *J. Signal Inf. Process.*, vol. 04, no. 03, pp. 173–175, 2013, doi: 10.4236/jsip.2013.43B031.

V. Klausner *et al.*, “Principal component analysis in the modeling of HILDCAAs during the Solar Minimum of Cycle 23/24,” *J. Atmos. Solar-Terrestrial Phys.*, vol. 213, p. 105516, Feb. 2021, doi: 10.1016/j.jastp.2020.105516.

W. Younas, C. Amory-Mazaudier, M. Khan, and R. Fleury, “Ionospheric and Magnetic Signatures of a Space Weather Event on 25–29 August 2018: CME and HSSWs,” *J. Geophys. Res. Sp. Phys.*, vol. 125, no. 8, pp. 1–18, Aug. 2020, doi: 10.1029/2020JA027981.

X. Wei and H. Ouyang, “Carbon price prediction based on a scaled PCA approach,” *PLoS One*, vol. 19, no. 1 January, pp. 1–16, 2024, doi: 10.1371/journal.pone.0296105.

Y. Zhang, Y. Wang, Y. Zhu, L. Yang, L. Ge, and C. Luo, “Visibility Prediction Based on Machine Learning Algorithms,” *Atmosphere (Basel)*, vol. 13, no. 7, p. 1125, Jul. 2022, doi: 10.3390/atmos13071125.

How to cite this paper:

Aznilinda Zainuddin, Muhammad Asraf Hairuddin, Zatul Iffah Abd Latiff, Nur Dalila Khirul Ashar, Anwar Santoso, Mohamad Huzaimy Jusoh, Ahmad Ihsan Mohd Yassin (2025). Feature Reduction on Sym-H Index Image Using Principal Component Analysis Approach. *Malaysian Journal of Applied Sciences*, 10(1), 65-75.