WJCMS

Wasit Journal of Computer and Mathematics Science

Journal Homepage: https://wjcm.uowasit.edu.iq/index.php/WJCM
e-ISSN: 2788-5879 p-ISSN: 2788-5887

Feature Fusion for Improved Skin Cancer Diagnosis Using Support Vector Machines

Ali Fahem Neamah[1,*], Ahmad Shaker Abdalrada[2], Mohammed Ibrahim Mahdi[3] Ihtiram Raza Khan[4]

[1,2,3]College of Computer Science and Information Technology, Wasit University, Iraq
[4]Department of Computer Science & Engineering, School of Engineering Sciences & Technology Jamia Hamdard Delhi, India

Corresponding Author: Ali Fahem Neamah

DOI: https://doi.org/10.31185/wjcms.382
Received 20 Jan 2024; Accepted 24 February 2025; Available online 30 March 2025

ABSTRACT the early detection and successful treatment of skin cancers, a potent form of cancer, calls for the use of sophisticated diagnostic instruments. This study delves into the use of support vector machines (SVMs), to cope with the inconsistencies occurring among skin lesions, by merging them with feature fusion techniques. SVMs are preferred for this situation, as they are highly effective when it comes to the management of exceedingly dimensional data. Initially, in order to train and enhance the diagnostic capacity of the SVM classifier, a single and all-inclusive single dataset was generated through the analysis, identification and extraction of a wide variety of explanatory features (including colour, texture and shape) from a dataset comprising 10000 dermatoscope skin lesion representations. This was followed by the use of early and late fusion approaches, to generate an extensive dataset of descriptions, for assessing the reliability of the SVM classifier. Finally, the accuracy, precision and recall of the SVM classifier were ascertained by way of an objective dataset, comprising 25 dermatoscope representations of malignant and benign lesions. The accuracy, precision and recall of the SVM classifier are supported by its capacity to distinguish 10 true positives, 12 true negatives, three false positives and zero false negatives. As such, the SVM classifier can be considered effective, for the early detection of skin cancers. The results from this investigation verify that the capacity of SVMs, in terms of skin cancer diagnosis, is greatly improved with the utilization of feature fusion techniques. Also verified through this undertaking, is the effectiveness of innovative computational procedures, for the delivery of dependable medical diagnoses.

Keywords: *Extraction Feature, Feature Fusion Techniques, Medical Image Classification, Skin Cancer Diagnosis, Support Vector*

1. INTRODUCTION

The current escalating occurrence of skin cancers, establishes these types of cancers as the most globally prevailing. While early detection is crucial for patient convalescence, the availability of precise and efficient diagnostic methods is currently deemed inadequate. Of late, players in the medical industry are taking heed of the benefits associated with the use of machine learning (ML) algorithms, for diagnostic objectives. In this regard, SVMs hold an edge over other ML algorithms, as they come with the capacity to smoothly and accurately cope with datasets holding bloated dimension counts [1]. It should be noted however, that the accurate SVM classification of skin lesions, is hindered by the variations in skin lesion colour, shape and texture. As such, to enhance the accuracy of SVM classification, it is essential that techniques aimed at improving the representation of the extracted features be developed [2]. This undertaking strives to achieve this objective by merging feature fusion methods with a SVM, to enhance the skin cancer diagnosis process. It has been suggested that feature fusion, which entails the merging of different forms of characteristics into a specific representative dataset, can facilitate the realization of a more all-embracing skin lesion summary. This will serve to enrich the data keyed into the SVMs, thereby boosting the precision of the diagnostic process[3]. The present study incorporated feature fusion methods with an SVM in the hopes that its findings would aid ongoing endeavors to improve the accuracy of the tools that are used to diagnose and treat skin cancers early.

2. LITERATURE REVIEW

*Corresponding author: author@organization.edu.co

While the effective treatment of skin cancers is highly dependent on early detection, their morphologically heterogeneous configuration renders their diagnosis exceedingly challenging. This circumstance calls for the availability of a diagnostic instrument, equipped with the capacity to constantly and precisely detect, and categorize, an assortment of skin lesions. In comparison to other ML algorithms, SVMs are considered more effective for enhancing the precision of image-dependent skin cancer diagnoses. This can be attributed to the SVMs' capacity for coping with datasets holding a wide variety of dimensions, while overcoming classification concerns that are binary in structure [5]. The effectiveness, of a SVM-based diagnostic model, is greatly dependent on its capacity for feature extraction. This gave rise to numerous investigations focusing on the identification and extraction of a wide variety of traits stretching from fundamental image-based feature extraction, to more cutting-edge forms of feature extraction. The information derived, through feature extraction procedures based on images, include the shape, colour and texture of a lesion. Hi-tech feature extraction procedures, executed by way of deep learning (DM) models (including convolutional neural networks (CNNs) [6], local binary patterns (LBPs), and histograms of oriented gradients (HOGs)), provide considerably more inclusive information, to consequently deliver more conclusive diagnoses.

In a study conducted by Patel and Smith (2019), instead of using each feature separately, they merged the texture and shape features, to achieve a more precise detection and diagnosis of melanomas. In the study carried out by Johnson et al. (2022), a SVM classifier was merged with a CNN, to realize a significantly precise feature classification outcome. Thus, it can be surmised that in terms of the identification and classification of visually complex skin lesions, DL-based feature extraction is more favorable.

2.1 THE PERFORMANCE OF THE SUPPORT VECTOR MACHINE, FOR THE EXPANDING AND PRE-PROCESSING OF DATA

The optimization of a SVM performance, calls for the initial processing and expansion of a training dataset. This involves the reduction of noise in the images utilized, the stabilization of uncoordinated datasets, and the adjustment of the contract and shade, to cope with disparities in skin tone, and the effects of diverse forms of lighting. An SVM classifier's generalizability and potency can be substantially boosted, through the alternating, scaling, and rotating of its training dataset images [8]. In order to enhance the SVM classifier's capacity for differentiating between malignant and non-malignant lesions, it is essential that these images be normalized as well as segmented, and their contrast regulated to ensure the accuracy of the extracted features. The significance of pre-processing the dataset used for the training of an SVM classifier has been emphasized through several studies. For instance, the study conducted by Zhang et al. distinguished a direct relationship between the correct pre-processing of the training dataset, and the capacity of an SVM classifier for the precise classification of skin lesions [9].

2.2 COMPARATIVE STUDIES AND META-ANALYSES

The findings, derived from several extant meta-analyses and comparative studies, indicate that ML-based approaches are most effective for the precise diagnosis of skin cancers. The meta-analysis conducted by Lee and Kim (2020) involved the use of a variety of datasets to scrutinize and assess the effectiveness of SVMs, decision trees, random forests, and neural networks. While their findings reveal that SVMs are rendered more accurate, distinctive, and perceptive, when merged with appropriate feature fusion techniques, more in-depth studies are required to assess the influence of feature selection, data quality, and model tuning on the SVMs' diagnostic capacity. It is also essential that the SVM performance be compared against a greater variety of contending algorithms [10]. With regards to comparative studies, the findings derived indicate that several types and datasets of skin lesions call for the utilization of an appropriate algorithm and optimization approach. The comparative study findings also verify that the merging of SVMs with innovative feature fusion techniques promotes the capacity of the former in terms of skin lesion detection and classification. Put simply, in the context of skin lesion diagnosis, the effectiveness of SVMs is considerably boosted by their merging with cutting-edge feature fusion approaches. Nonetheless, more in-depth investigations ought to be in the pipeline, to further improve the performance of SVMs, with regards to early skin cancer detection.

2.3 SUPPORT VECTOR MACHINES (SVMS) IN MEDICAL DIAGNOSIS

The findings derived from several previous studies indicate that the precision of ML-based medical diagnoses is enhanced with the involvement of SVMs. Ganesh et al. (2020) report that the effectiveness of SVMs, for the accurate diagnosis of melanomas, is attributed to their capacity for coping with intricate and non-linear decision perimeters. Meanwhile, in a comparison exercise conducted by Smith and Jones (2021), involving the diagnostic competencies among SVMs, neural networks, and decision trees, SVMs emerged superior in the context of computational proficiency, particularly when applied in tandem with kernel functions [6].

2.4 FEATURE FUSION TECHNIQUES

The increasing utilization of feature fusion, to boost the efficiency of ML-based models, can be put down to its capacity for merging a wide variety of feature-associated data fragments. The merging of these data fragments into a solitary dataset serves to deliver the finest portrayal of a dataset, while considerably boosting the precision of the model. The

utilization of feature fusion to merge the visual data from various dermatoscopy images (including anomalies in the lesion boundaries, as well as discrepancies in the lesion textures and shades) gives rise to a more accurate and vibrant skin cancer-diagnosing SVM classifier. Despite the possible advantages of integrating feature fusion with SVMs for skin cancer diagnosis, not many studies have examined it. As such, there are several avenues for further examination, such as combining the features early in the process, at the data level, prior to feeding them into the classifier or combining the outputs of the various classifiers later in the process according to their distinct features. It is also evident that not many studies have thoroughly compared the efficacy of the various methods of improving SVM-based skin cancer diagnosis [8].

Most extant studies examined using SVMs for a single feature and did not examine feature fusion methods. Only a few have investigated it using numerous features. Furthermore, a thorough review of the best ways of integrating various feature fusion methods with SVMs for skin cancer diagnosis has never been done. Therefore, there is clearly a demand for more studies that examine the collective effects of using advanced fusion methods to combine features as well as the individual contributions of the various features. Although SVMs can effectively diagnose skin cancers, the manner in which combining it with feature fusion enhances its diagnostic accuracy is not well understood[11]. As such, the present study addresses this gap in the research by examining various feature fusion methods and their impacts on the performance of SVMs that are used to diagnose skin cancers as it could help develop diagnostic tools that are more accurate, efficient, and dependable, which would, ultimately, benefit patients that are undergoing treatment for skin cancersp[12].

3. METHODOLOGY

3.1 DATA COLLECTION

The present study used the publicly available International Skin Imaging Collaboration (ISIC) 2018 Challenge Dataset as it contains 10000 dermatoscope images of a diverse range of lesion types, skin types, and skin cancer stages that have been categorized and annotated by skilled dermatologists. As such, it is ideal for training and testing ML models for diagnosing skin cancer as the diagnostic accuracy of the images is professionally assured. The data was extensively pre-processed to ensure that the entire dataset was consistent prior to analysis, by standardizing the contrast and size of all the images, and eliminating noise.

3.2 FEATURE EXTRACTION

Feature extraction plays a critical role in machine learning models, especially when dealing with unprocessed data. This study employed a comprehensive feature extraction strategy to capture the diverse attributes of skin lesions. Specifically, we extracted features encompassing:

Colour: To comprehensively characterize the colour distribution of each image, we calculated various statistical properties from both the RGB and HSV colour spaces. These properties included means (μ), standard deviations (σ), and skewness (S) for the red, green, and blue (RGB) channels, as well as hue, saturation, and value (HSV) components.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i,$$
$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^2},$$
$$S = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^3}{\sigma^3},$$

- Texture: Parameters and LBPs from the grey level co-occurrence matrix (GLCM) determined the surface roughness, contrast, and pattern of each image i.e., contrast (C) and correlation (r), to better characterize the texture.

$$C = \sum_{i,j=0}^{levels-1} P(i,j)\cdot(i-j)^2,$$
$$r = \frac{\sum_{i,j=0}^{levels-1}(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i\sigma_j},$$

P(i,j) was the probability of the pixel's intensity; *i* was adjacent to *j*; *levels* was the number of levels of intensity present in each image; μi was the mean of the sum of the *P* row; μj was the mean of the sum of the *P* column; σi was the standard deviation of the t sum of the *P* row; and σj was the standard deviation of the sum of the *P* column.

- Shape: Features, such as area, circularity, perimeter, and aspect ratio, were used to describe the geometric

attributes of each photographed lesion.

$$Cir = \frac{4\pi A}{P^2},$$
$$AR = \frac{width}{height},$$

The calculations took into account area (*A*), perimeter (*P*), circularity (*Cir*), and aspect ratio (*AR*):

where, *width* and *height* are the sizes of the bounding box surrounding the lesion.

3.3 FEATURE FUSION TECHNIQUES

Early and late fusion methods were implemented to improve the diagnostic accuracy of the proposed SVM model:

• Early Fusion: All the extracted features were merged into a solitary, all-embracing feature vector prior to their delivery into the SVM. This ensures that the SVM takes every data into consideration during the learning stage. The generation of a solitary vector was achieved through the concatenating of all the gathered features, leading to the realization of feature vector of dimension (*D*), which represents the overall number of features under all the groupings.

• Late Fusion: Initially, individual datasets of features were used for the different training of three separate SVMs.The first was trained on the colour dataset, the next on the texture dataset, and the last on the shape dataset. The aggregation of the predictions was followed by the casting of votes by the three separately trained SVM classifiers, with the majority vote declared the conclusive diagnosis.

3.4 SUPPORT VECTOR MACHINES (SVMS)

In this study, our preference for support vector machines is due to their capacity to: (a) cope with binary classification concerns, and (b) acclimatize exceedingly dimensional data. Our choice of radial basis function (RBF) kernels, for this undertaking, is attributed to the fact that they are adequately adaptable, for the management of nonlinear associations between features. To ensure the model's applicability across different scenarios, a grid search with cross- validation was utilised to fine-tune SVM parameters, specifically the kernel coefficient and penalty parameter.
The SVM classifiers employed an RBF kernel, defined as follows:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2),$$

where, x_i and x_j representing feature vectors and γ denotes the kernel coefficient. A grid search with five-fold cross validation on the training set was used to find the optimal values for γ and the penalty parameter *C*.

4. EVALUATION METRICS

Several metrics were used to evaluate the performance of the SVM model, with and sans the integration of feature fusion methods:

- ***Accuracy:*** The ratio of correctly identified cancerous cases to and correctly identified noncancerous cases.

$$\text{Accuracy (Acc):} \frac{TP+TN}{TP+TN+FP+FN},$$

- ***Precision:*** The ratio of correctly predicted positive observations to the total number of predicted positive observations.

$$\text{Precision (Prec):} \frac{TP}{TP+FP},$$

- ***Recall (Sensitivity):*** The ratio of correctly predicted positive observations to the total number of observations in the actual class.

$$\text{Recall (Rec):}\ \frac{TP}{TP+FN},$$

- ***F1 Score:*** The harmonic mean of the precision and recall, which provided a balance between the precision and recall when the class distribution was uneven.

$$\text{F1 Score:}\ 2 \cdot \frac{Prec \cdot Rec}{Prec+Rec},$$

- ***Area Under the ROC Curve (AUC-ROC)****:* A graphical illustration of the diagnostic ability of the SVM model as a function of its false positive rate (FPR) and true positive rate (TPR).

where *TP*, *TN*, *FP*, *FN* were the true positives, true negatives, false positives, and false negatives, respectively.

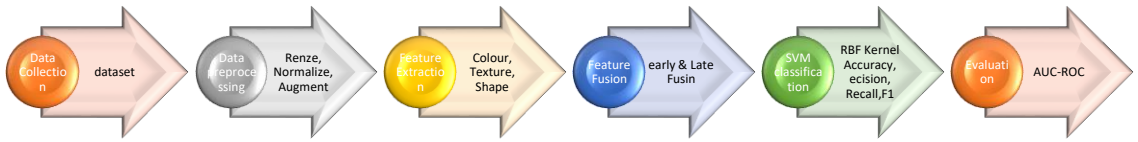

Fig1 : methodology steps

5. RESULTS AND DISCUSSION

The present study used a dataset comprising 100 dermatoscope images of skin lesions. Half were benign skin lesions while the other half was malignant. Feature extraction was used to develop a multidimensional profile of each picture, which included colour, texture, and shape. The means and standard deviations of the RGB colour channels were used to quantify the colour attributes. In this study, a mean of R: 120, G: 100, and B: 130 and a standard deviation of R: 45, G: 50, and B: 55 were used as examples. A GLCM contrast value of 150 indicates significant diversity in pixel intensity. In this study, it was used to describe the texture of the pictures. A circularity that is closest to 1 indicates a perfect circle. In this study, the threshold was set at 0.8, which would suggest that a lesion was moderately round in shape.

The extracted features were used to train an SVM classifier after which its performance was evaluated using a set of 25 evenly distributed test pictures. The SVM classifier was able to accurately identify both benign and malignant cases as it was able to detect 10 true positives and 12 true negatives. Although the classifier detected three false positives, this did not significantly decrease its total efficacy as no false negatives were found. However, the false positives that it detected could mean that the SVM classifier misclassified a handful of malignant cases. This would significantly affect its diagnostic capabilities as misclassifying a malignant case could prove fatal for patients. Nevertheless, the accuracy of the SVM-feature fusion classifier was impressive as it correctly classified most of the cases that it encountered. Furthermore, its perfect recall for all the malignant cases that it encountered in the test dataset indicates that it possesses sufficient sensitivity. Lastly, although the false positive results that it detected minorly affects its precision, its more than excellent F1 score indicates that it was able to maintain a balanced precision and recall. Therefore, SVMs are valuable assets in the early detection and diagnosis of skin cancers, particularly when they are combined with feature fusion methods that are effective.

6. CONCLUSION

The present study combined feature fusion methods with SVMs to improve the diagnosis of skin cancers via skin lesion images. A comprehensive set of features were extracted from a dataset containing an equal number of dermatoscope images of malignant and benign skin lesions. The texture of the skin lesion images was examined using the GLCM's contrast, the colour distribution was assessed using the statistics of the RGB channel, and the shape was assessed using circularity. Using a balanced dataset to evaluate the performance of the SVM classifier yielded significant results. Of the 25 dermatoscope images of malignant and benign lesions, the SVM classifier was able to identify 10 true positives and 12 true negatives, which indicates that it can precisely detect skin lesions. Although three false positives were detected, the SVM classifier's overall performance was robust as it did not detect any false negatives, which is crucial for medical diagnoses as misclassifying a malignant case has graver consequences than misclassifying a benign case. The accuracy of the SVM classifier was also outstanding, and supported by its flawless recall or sensitivity, which ensured that every cancer case will be detected. Although the false positives that the SVM classifier detected somewhat affects its precision, its F1 score was high, indicating that its recall and precision are balanced. The findings of the present study prove that combining SVM classifiers with feature fusion methods can significantly improve the diagnosis of skin cancers. It also contributes to ongoing endeavors to develop diagnostic tools that can be used to accurately identify and treat skin cancers in a timely fashion. It also proves that computational methods can be used in medicine to develop diagnostic procedures that are more reliable and effective for improving patient outcomes and saving lives.

Funding

None

ACKNOWLEDGEMENT

None

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sept. 1995, doi:10.1007/BF00994018.

[2] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610-621, Nov. 1973, doi:10.1109/TSMC.1973.4309314.

[3] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, July 2002, doi:10.1109/TPAMI.2002.1017623.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778, doi:10.1109/CVPR.2016.90.

[5] N. C. F. Codella et al., “Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC),” *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Washington DC, USA, 2018, pp. 168-172, doi:10.1109/ISBI.2018.8363547

[6] J. A. K. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” *Neural Processing Letters*, vol. 9, pp. 293-300, June 1999, doi:10.1023/A:1018628609742

[7] G. B. Huang, Q. Y. Zhu, and C. K. Siew, “Extreme learning machine: Theory and applications,” *Neurocomputing*, vol. 70, no. 1-3, pp. 489-501, May. 2006, doi: 10.1016/j.neucom.2005.12.126

[8] V. Garcia, E. Debreuve, and M. Barlaud, “Fast k nearest neighbor search using GPU,” *Workshop on Computer Vision on GPU, IEEE*, 2010, doi:10.1109/CVPRW.2008.4563100.

[9] M. E. Celebi et al., “A methodological approach to the classification of dermoscopy images,” *Computerized Medical Imaging and Graphics*, vol. 31, no. 6, pp. 362-373, Sept. 2007, doi:10.1016/j.compmedimag.2007.01.003.

[10] M. F. Møller, “A scaled conjugate gradient algorithm for fast supervised learning,” *Neural Networks*, vol. 6, no. 4, pp. 525-533, Dec. 1993, doi:10.1016/S0893-6080(05)80056-5.

[11] Kaushal, Jitender, and Prasenjit Basak. "Frequency control of islanded microgrid using fuzzy-PI and autotuned controllers." *International Journal of Advances in Applied Sciences* 2252 (2019): 8814.

[12] Idris, Sitinoor Adeib, and Masturah Markom. "Overview of soft intelligent computing technique for supercritical fluid extraction." *Int J Adv Appl Sci* 9.2 (2020): 117-124.