B. Arivuselvam[1], S. Tanisha[2], S. Shalini[3], V.S. Subhalaksmi[4]

Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12,Issue 8 July 2021:5595–5606

Research Article

SKIN CANCER DETECTION AND CLASSIFICATION USING SVM CLASSIFIER

B. Arivuselvam[1], S. Tanisha[2], S. Shalini[3], V.S. Subhalaksmi[4]

Abstract

Human cancer is the most hazardous sicknesses existing which is principally brought about by hereditary flimsiness of numerous atomic changes. Among the numerous kinds of disease, skin cancer is quite possibly the most widely recognized sorts of malignancy. There are three kinds of skin malignant growth, to be specific, Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC)and Melanoma, melanoma is the sort of skin cancer which is perilous. The skin cancer detection technology is extensively isolated into four fundamental parts beginning from gathering dermoscopic image data set, dermoscopic image database, image pre-processing which includes hair removal, noise removal, sharpening, resize, contrast stretching of the given skin image, segmentation in which gave for segmenting the zone of interest from the given image. Various methods can be utilized for segmentation. Some regularly utilized division calculations are k-means, thresholding histogram and so on, feature extraction from the portioned picture and grouping of the picture from the feature set separated from sectioned picture. Various classification algorithms are used for this, among which the utilization of machine learning and deep learning-based algorithm are used to improve results for classification. The most frequently utilized classification. algorithms are 'support vector machine', 'feedforward artificial neural network', 'deep convolutional neural network'. This paper provides the two types of skin cancer - Basal Cell Carcinoma and Melanoma and equally threatening (skin) diseases such as Actinic keratosis, Cherry nevus, Dermatofibroma and Melanocytic nevus, and classify them into six different classes using the 'support vector machine (SVM) classifier'.

Keywords- *skin cancer, skin disease, Support Vector Machine (SVM)classifier, GLCM, Gabor filter, Fuzzy Clustering*

INTRODUCTION

Skin cancer is the unusual development of skin cells which frequently develops on skin exposed to the sun. However, having referenced that, it might likewise happen on territories of the skin not exposed to the daylight. Among three sorts of skin disease, such as, Squamous Cell

[1]Assistant Professor, Dept. of ECE, Easwari Engineering College
[2, 3,4]Students, Dept of ECE, Easwari Engineering College

Carcinoma (SCC), Melanoma and Basal Cell Carcinoma (BCC), Melanoma is generally perilous in which endurance rate is extremely low. The other two goes under the non-melanoma malignancy. As indicated by the American cancer society's evaluations for melanoma in the USA in 2021, around 106,110 instances of melanoma will be analyzed. Among them, 7180 are expected to die as a result of melanoma which is for the most part brought about by the skin's openness to the sun. From a study done by a UK College, it is tracked down that 86% of melanomas are uncovered by bright (UV) radiation. Overall, individuals' danger for melanoma copies on the off chance that the person has had in excess of five burns from the sun. In spite of the fact that melanoma is a hazardous illness around the world, early discovery of melanoma alongside other skin cancer types can build the opportunity of endurance of the person in question. In the event that the individual creates dubious spots or developments on the skin, or notification any progressions in existing spots or developments, the specialist will inspect the skin or alludes to an expert for determination. The expert will look at the shape, size, surface and shade of the dubious region on the skin. In the event that it is suspected to be dangerous, a skin biopsy method might be performed, in which they will eliminate the dubious zone or a part of it to ship off a lab for testing, which is costly and incredibly excruciating and requires a long time to think of the outcome. Consequently, there is aincreasing requirement for programmed skin disease discovery framework with high exactness. The identification of the skin malignant growth consistently accompanies the cycle of biopsy that should be performed by the particular dermatologist. There are two types of picture accessible for skin disease recognition. The dermoscopy picture is caught by specific devoted framework in neurotic community with center around area of concern with high zoom, which needs talented dermatologist to finish up the picture as sure or negative. This kind of picture can be taken care of to modernized semiautomated framework for characterization. [2]. Yet, in this development, the loss for each situation needs to walk around the over-the-top local area and prerequisites to take consultancy of the dermatologist. Of course, if there is a PC programming which can normally recognize skin malignancy from modernized picture got by any high-level picture getting structure with little focus on the area of concern, setback can at whatever point play out the test and later request the expert assessment in the event of the positive consequence of the skin disease. Subsequently the inspiration of this paper is to comprehend the new technology for skin cancer and skin disease discovery and spotlight on building up a programmed framework for skin cancer growth and skin disease location and characterizing them into various classes from computerized picture utilizing machine technology.

RELATED WORKS

Skin cancer detection utilizing non-obtrusive strategies, for example, picture preparing procedures got one of the appealing and requesting research in the recent few years.Wiltgen, utilizes a strategy for tissue counter examination (TCA), that depends on partitioning the entire picture into square components of equivalent size and afterward includes are determined from these square components of the picture. The features depend on GLCM (Gray level co-event lattice) and grey level histogram where the separation of homogeneous and high difference territories happens.

PrasitthichaiNaronglerdrit, a design for melanoma location from dermatoscopic pictures was introduced. The design comprises of picture pre-processing for skin hair removal, the zone of

interest extraction or the skin sore, picture standardization and characterization utilizing re-trained deep convolutional neural networks with transfer learning.

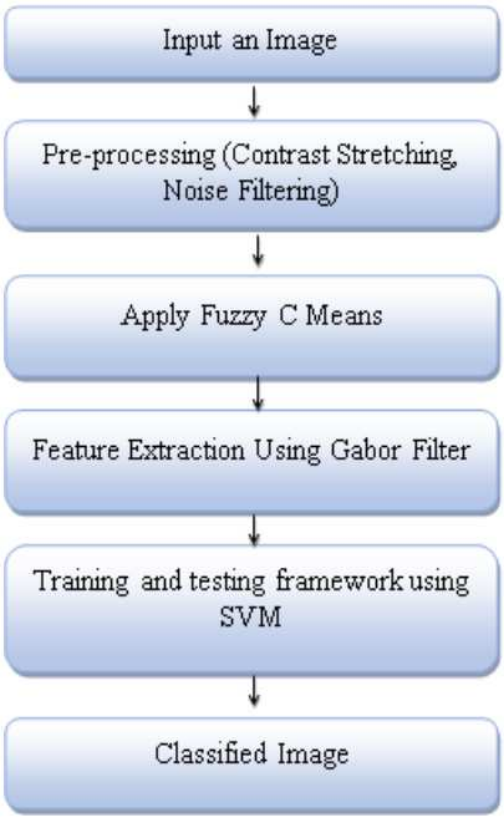
For the segmentation process, existing frameworks either utilize manual techniques or programmed border recognition strategies. In numerous papers, the highlights like shape, color, texture and luminance are utilized for isolating the skin lesions during the segmentation process.

PROPOSED METHODOLOGY AND PROCEDURE

For Image Acquisition, the medical images are acquired from a reliable and trusted source. The software utilized for this is MATLAB, which is installed, and the image acquisition, pre-processing and segmentation process are done. Pre-processing – Noise found in images are filtered by replacing the value of each pixel with average of pixels around it and image is enhanced through contrast stretching. Segmentation- Fuzzy C(clustering) algorithm is used. Feature extraction- Done by GLCM and Gabor filter which extracts features like texture, colour, size from the input images.

Classification: SVM classifier will calculate the feature values of test and database images and classify accordingly. The dataset is trained with 1500 images which consists of images from the four types of skin diseases and the two types of skin cancer mentioned.

FLOWCHART



SKIN CANCER IMAGE ACQUISITION

In this segment, the specifics of freely accessible skin cancer picture information bases utilized are talked about. For, the identification of skin cancer principally two sorts of pictures utilized aredermoscopy picture and digital picture. The dermatoscopic pictures are caught by particular committed framework in obsessive place which is centeredaround zone of interest with high zoom. Then again, computerized pictures are caught by any advanced picture catching framework with little spotlight on the area of concern. Dermoscopy pictures are more appropriate for conclusion, yet they need master dermatologist to distinguish skin malignant growth. [2],[5]. There is appeal for skin cancer detection technology based on digital image. The dataset utilized here are gathered from different trusted and legitimate sites like Medicine.Dermweb, Dermis, Dermquest and from Kaggle. The absolute number of pictures utilized in the dataset are 1500. All these are dermatoscopic images. This dataset has each of the six classes images of different resolutions.

IMAGE PRE-PROCESSING

Images gathered for skin disease recognition are not reasonable for direct use of classification algorithm. Skin pictures may contain diverse undesirable antiquities like fine hair, air bubbles, and noise. Noise must be reduced from all images in the dataset.

1) Noise reduction

Median filter is one of the types of filters which is used for reducing noise from the skin images. Image noise is frequently created by the picture sensor and hardware of a scanner or camera. It is an undesirable by-product of image catch that impedes the predetermined data. In this, we use the Image processing toolbox is the one which provides a set of standard reference algorithms that can perform image segmentation, image enhancement, noise reduction, etc.

2) Contrast Enhancement

Contrast enhancement is a procedure to improve the representation of pictures. For skin cancer pictures, it is once in a while important to upgrade the Contrast in the area of interest. Histogrambalance is quite possibly the most usually utilized procedures for Contrast enhancement. Histogram evening out consistently appropriate the pixel force across the pictures in view of which pictures have more improved perception and utilizations a basic histogram equalizer for contrast upgrade.

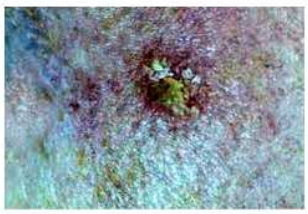
Melanoma image

Actinic Keratosis image

Pre-processed image-

Melanoma

Actinic Keratosis



C. IMAGE SEGMENTATION

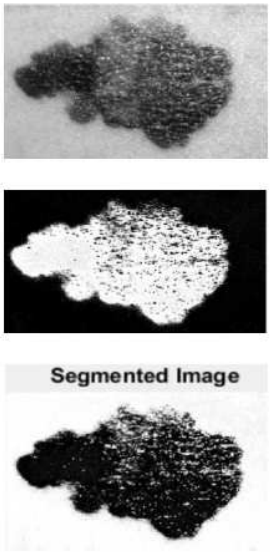
The next stage is image segmentation, which is a technique used in image processing to separate an image into numerous parts, depending on the qualities of the pixels in an image. Image segmentation includes detecting and segmenting the region of interest, based on how identical they are in color or shape. After the pre-processing technique is applied to an image which makes it free from all kinds of impurities, we need to segment that image focusing on the region of interest, to make the classification easy. An image segmentation technique called clustering, is an approach which is used to separate a group of elements in a region.

a. Fuzzy C (Clustering)

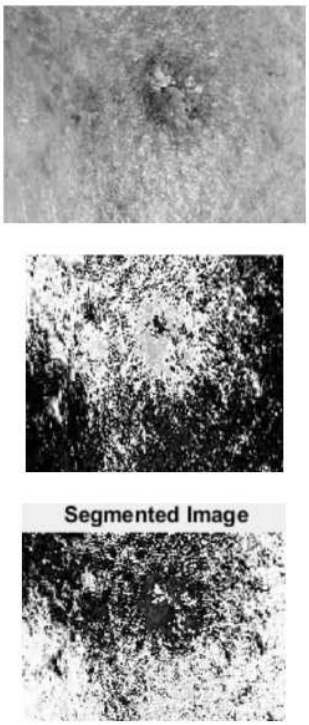
In In our paper we have utilized Fuzzy C- means (FCM) algorithm for the segmentation process. Fuzzy C-means is an unsupervised clustering algorithm. Its logic is derived from fuzzy set theory. We have gone for FCM rather than k-means clustering because the final outcome is clear and less complex, and separation of elements is done almost perfectly. The reason is because in k-means algorithm, the data point can be included in only one cluster making it harder to separate but in FCM, the data point can be included in more than one cluster or multiple

clusters. Here, the data points lying closest to the centre of a cluster will have a high priority in that cluster, and the data points lying farther away from the centre of a cluster will have a lower priority in that cluster. Now with the image segmented, it makes it easier to extract the texture and color features. [12-16].

The stages in segmentation-Melanoma



Actinic Keratosis



D. Feature extraction

Feature extraction is a process where various features are extracted from an image and converted into numerical data while the original information remains unchanged in the dataset. After the region of interest is partitioned in the image from the previous process, we move onto feature extraction where many features like texture, color, diameter, size etc are extracted which is later useful for the classification process.

In this paper, two main features are extracted to help differentiate between each of the six classes- color and texture. The texture analysis is done with the help of the GLCM (Gray Level Co-occurrence Matrix) features and Gabor Filter, and the texture element is extracted from RGB colored picture.

1) Gray-Level Co-occurrence Matrix

The GLCM is utilized for analyzing the texture feature. Here, the spatial relation that each pixel has with other pixels is considered for viewing the texture of an image. They compute how often a pair of pixels with certain values and in a spatial relation arise in an image thereby forming a matrix and obtaining information from them. They assist in characterizing the texture present in an image. A gray co-matrix function is used to create a GLCM. The following statistics give some details about the texture extracted from an image.

Contrast- measures intensity between two pixels. We can measure it using the formula-

$$\sum_{i,j} |i-j|^2 p(i,j) \qquad (1)$$

Correlation-is the joint correlation or probability of any two pixels in an image.

$$\sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j} \qquad (2)$$

Energy-is the addition of the squared components in the Gray Level Co-occurrence Matrix.

$$\sum_{i,j} p(i,j)^2 \qquad (3)$$

Homogeneity- it shows how close the distributed objects are present in the Gray Level Co-occurrence Matrix.

$$\sum_{i,j} \frac{p(i,j)}{1+|i-j|} \qquad (4)$$

Here i and j are the first and last values in the range,

P represents the probability between two pixels,

μ represents the mean,

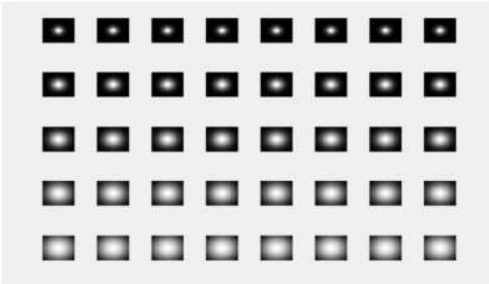
σ represents the standard deviation.

Other GLCM features used are smoothness, skewness, mean, standard deviation, variance, rms, entropy, kurtosis and Inverse Difference Moment (IDM).

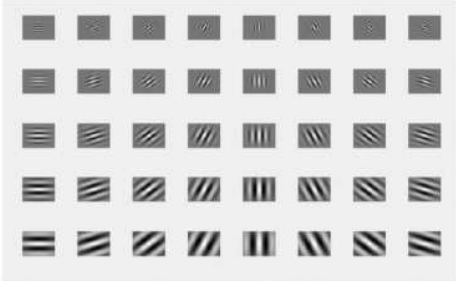
2) Gabor filter

We have opted for Gabor filter as they are the filters which are mostly utilized for the analysis of texture. They are linear filters and perform this by focusing on the region of interest and analyse if there is any frequency content around it. Most people use this as a model for recognizing texture because the methodology of this filter is similar to how people differentiate texture usually. The Gabor filter has a Gaussian function with variance values and centre frequencies [13]. In MATLAB, through the image processing toolbox, the function "imgaborfilt" is used for applying Gabor filter to an image.

Magnitude (Mag) of the Gabor Filter in the given image



Real part (Re) of Gabor Filter on the given image



Color features-The analysis of the texture features is done with the help of the Gabor Filter. Now, color plays a significant part in diagnosing skin cancer. Dermatologists believe that a possible symptom of skin cancer could be the variation of color of the affected part in the skin region. To extract the color features, the statistical parameters like variance, entropy, mean and standard deviation are determined.

The main GLCM features were computed as follow.

Class	Contrast	Cor-relation	Energy	Homogeneity
Actinic keratosis	0.0915	0.1345	0.0968	0.1362
Cherry nevus	0.0534	0.0672	0.0411	0.0636
Dermatofibroma	0.0396	0.0417	0.0306	0.0404
Melanocytic nevus	0.0101	0.0101	0.3160	0.0303

Basel cell carcinoma	0.1013	0.1190	0.0385	0.1200
melanoma	0.0791	0.002	0.0095	0.0100

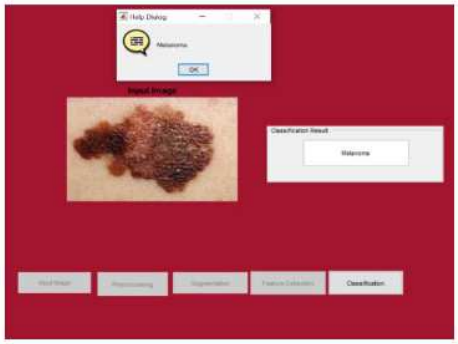
E. Image Classification

In our paper, for image classification we have used the SVM classifier. SVM is a MLalgorithm for data classification. SVM's core idea is decision planes that is used to give information about the decision boundaries. A decision plane separates objects into classes by separating them. SVM employs a supervised ML algorithm in which each and everydata isplotted in a space of n dimensions such that the n represents total features and the value of these features is obtained from specific co-ordinate. By maximising the separating hyper plane inbetween data points of the training dataset, SVM attempts to assign the data points to a class. SVM attempts to increase the distance of both hyperplanes from the separating hyperplane during training, and thus classification is completed by figuring the hyperplane that separates the classes [3], [5], [6].

The dataset is trained with 1500 images belonging to the six classes and with the help of GLCM the texture and color features extracted from each of the six classes are given as input to the SVM. The SVM makes use of supervised algorithm utilizing the element vectors which are the color and texture vectors in our case to prepare and train the database. Each image's features, colour, and texture are saved in the database and will be used in the further stages of classification. Distance metrics are also employed to measure similarities of features among images for better classification. In this case, we will square measure practise the GLCM selections for detection. Skin lesions square measure extracted normalised symmetrical grey Level Co-occurrence Matrices GLCM Texture options using GLCM Each of the four classes' square measure is extracted and fed into the Multi-Class Support Vector Machine, which is used for classification. Similarity is analysed using Support Vector Machine classifiers among the features of the Test image and that of the database images in this case. The feature values of the input image, Test image, and database image is calculated by the SVM classifier and based on these values and how similar they are, the images are classified into each of these six classes accordingly. [6].

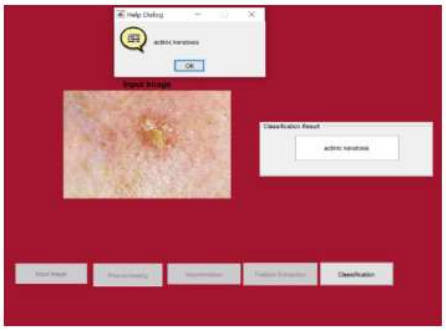
Classification result:

For Melanoma



SKIN CANCER DETECTION AND CLASSIFICATION USING SVM CLASSIFIER

For Actinic Keratosis



RESULTS AND DISCUSSION

a. Evaluation metrics

To quantitatively assess the performance of our proposed method certain parameters are used they are accuracy, sensitivity, precision and specificity

The proportion of correctly identified positive samples among all positive samples is measured by sensitivity.

Sensitivity = TP/ TP + FN (4)

The proportion of correctly identified negative samples among all negative samples is measured by specificity.

Specificity = TN/ TN + FP (5)

Precision is the degree to which estimates from different samples agree with one another.

Precision = TP/ TP+FP (6)

The overall rate of perfectly identified samples is referred to as accuracy.

Accuracy=TP+TN/ TP+FP+TN+FN(7)

Where,

TP= true positives

FP= false positives

TN= true negatives

FN= false negatives

Class	TP	FP	FN	TN
Actinic keratosis	244	65	56	1135
Cherry nevus	249	67	51	1133

Dermatofibroma	229	60	71	1140
Melanocytic nevus	255	48	45	1151
Basel cell carcinoma	235	53	65	1147
melanoma	230	65	70	1135

Based on these parameters, four evaluation metrics were calculated as follows

Class	Sensitivity	Specificity	Precision	Accuracy
Actinic keratosis	0.8133	0.9458	0.7896	0.9194
Cherry nevus	0.83	0.9441	0.7879	0.9214
Dermatofibroma	0.7633	0.95	0.7923	0.9226
Melanocytic nevus	0.85	0.9551	0.8415	0.9126
Basel cell carcinoma	0.7833	0.9558	0.8159	0.9213
melanoma	0.7666	0.9458	0.7796	0.91

The overall values of these evaluation metrics were found.

Metric	(%)
Accuracy	92.04
Sensitivity	80.11
Specificity	95.01
Precision	80.17

CONCLUSION

We were able to detect and classify not only the two skin cancer types (Melanoma and Basal cell carcinoma) in our study, but apart from that also four other skin diseases such as Actinic keratosis, Cherry nevus, Dermatofibroma and Melanocytic nevus. Since the mentioned skin diseases have similar symptoms to the two types of skin cancer at the beginning stage, this classification system helps in avoiding any confusion. The SVM classifier is also found to be more accurate and lowers the computational complexity. The results are also generated within five minutes which further helps in accelerating the process of getting the skin cancer treated if detected so.

For our future work, we would like to implement our system in mobile applications suitable for both android and iOS to make it more compactable.

REFERNCES

1. American Cancer Society. Cancer Statistics Center. http://cancerstatisticscenter.cancer.org. Accessed January 25, 2021
2. Suganya R. 'An automated computer aided diagnosis of skin lesions detection and classification for dermoscopyimages.'(2016) International Conference on Recent Trends in Information Technology (ICRTIT).
3. Farooq, M. A. Azhar, M. A. M. Raza. 'Automatic Lesion Detection System (ALDS) for Skin Cancer Classification Using SVM and Neural Classifiers.'(2016) IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE).

4. Joseph S,&Panicker J. R. 'Skin lesion analysis system for melanoma detection with an effective hair segmentation method.'(2016) International Conference on Information Science (ICIS).
5. N. Codella J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith. 'Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images.' (2015) in Proc. Int. Workshop Mach. Learn. Med. Imag. Heidelberg, Germany.
6. Uzmabano Ansari, Tanujasarode.'Skin cancer detection using SVM.'(2017) Proceedings of WRFER International Conference, Pune, India.
7. B. Harangi. 'Skin lesion classification with ensembles of deep convolutional neural networks.' (2018) J. Biomed. Informat.
8. A. M. Taqi, F. A. Azzo, A. Awad, and M. Milanova. 'Skin lesion detection by Android camera based on SSD-mobilenet and TensorFlow object detection API.' (2019) International J. Adv. Res.
9. Barata, C., & Marques, J. S. 'Deep Learning for Skin Cancer Diagnosis with Hierarchical Architectures.'(2019) IEEE 16th International Symposium on Biomedical Imaging (ISBI).
10. Hosny K. M, Kassem M. A., &Foaud M. M. 'Skin Cancer Classification using Deep Learning and Transfer Learning.'(2018) 9th Cairo International Biomedical Engineering Conference (CIBEC).
11. Nida, Nudrat&Irtaza, Aun&Javed, Ali &Yousaf, Muhammad Haroon & Mahmood, Muhammad. 'Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy C-means clustering.'(2019) International Journal of Medical Informatics.
12. Premana A., Wijaya A. P. &SoelemanM. A. 'Image segmentation using Gabor filter and K-means clustering method.'(2017) International Seminar on Application for Technology of Information and Communication.
13. Murugan, S., Jeyalaksshmi, S., Mahalakshmi, B., Suseendran, G., Jabeen, T. N., & Manikandan, R. (2020). Comparison of ACO and PSO algorithm using energy consumption and load balancing in emerging MANET and VANET infrastructure. *Journal of Critical Reviews*, *7*(9), 2020.
14. Subramaniyan, Murugan, et al. "Deep Learning Approach Using 3D-ImpCNN Classification for Coronavirus Disease." Artificial Intelligence and Machine Learning for COVID-19 (2021): 141-152.
15. Alzubi, J. A. (2021). Bipolar fully recurrent deep structured neural learning based attack detection for securing industrial sensor networks. *Transactions on Emerging Telecommunications Technologies*, *32*(7), e4069.
16. Alzubi, J. A., Jain, R., Kathuria, A., Khandelwal, A., Saxena, A., & Singh, A. (2020). Paraphrase identification using collaborative adversarial networks. *Journal of Intelligent & Fuzzy Systems*, *39*(1), 1021-1032.