Comparative Study of Binary Classifiers for Reducing False Negative Detection of Melanoma in Skin Lesions

Amith Jooravan
Department of Electronic and Computer Engineering
Durban University of Technology
Durban, South Africa
Amithj@dut.ac.za

Serendra Reddy
Department of Electronic and Computer Engineering
Durban University of Technology
Durban, South Africa
Serenr@gmail.com

Nelendran Pillay
Department of Electronic and Computer Engineering
Durban University of Technology
Durban, South Africa
Trevorpi@dut.ac.za

Abstract - **Reliable and accurate classification of a skin lesion is essential to the early diagnosis of skin cancer, especially melanoma. Traditional classification methods require performing a biopsy on the lesion. The overlap of benign and malignant clinical features may lead to incorrect melanoma diagnosis and/or excising an excessive number of benign lesions. This paper focuses on the use of machine learning to aid physicians with the non-invasive classification methodology of skin lesions, whilst prioritising the minimization of false negative classification. The clinical features used are based on the ABCD rule, representing the asymmetry, border, colour and diameter of the lesion. The dermoscopic images chosen are of melanoma lesions less than 0,76mm in thickness which corresponds to the early stages of cancer. The investigated classification methods include K-Nearest neighbours (KNN), Naïve Bayes and linear support vector machine. (LSVM). This research proposes the use of a LSVM machine learning algorithm to classify a skin lesion as being either melanoma or non-melanoma with the lowest false negative rate of the investigated classification. Classification accuracy of 85% and a false negative rate of 5% is achieved.**

Keywords - Skin cancer, Image processing, Support vector machine, Naïve Bayes, K-nearest neighbours

I. INTRODUCTION

South Africa is exposed to a high amount of ultraviolet (UV) radiation from the sun, owing to its geographical location. Increased exposure to UV radiation is considered one of the primary contributing factors to skin cancer among humans. Melanoma, which is one of the main types of skin cancer, is considered to be the most dangerous form [1]. If left untreated it can metastasize to other organs, severely reducing the patient survival rate. It has been shown that the survival rate of a patient significantly increases if melanoma is detected early [2].

In 1985, Dr. Robert Friedman developed the method for visually differentiating between a benign nevus and a malignant melanoma [3]. The clinical characteristics of early melanoma were referred to as the ABCD rule, representing the *asymmetry*, *border*, *colour* and *diameter* of the lesion.

Later, the characteristic *evolving* was added and subsequently amended to the ABCDE rule [4]; this is illustrated in Fig. 1.

According to the South African melanoma advisory board (SAMAB), the process for the diagnosis of skin cancer, such as melanoma, requires an initial visual inspection of the skin lesion by a dermatologist. Consequently, if the prognosis cannot rule out skin cancer, a biopsy by a pathologist is required to make a diagnosis [5].

To perform a biopsy, it is required that part or all of the suspicious mole be removed, by scraping or cutting, and sent to a pathologist for examination and analysis under a microscope. A subjective naked eye examination of a suspected lesion is approximately 60% accurate in detecting melanoma and up to 68% accurate using dermoscopic images [2, 6, 7].

Fig. 1. ABCDE rule [4]

This research attempts to provide a method to reduce the number of false negative prognoses by dermatologists. The classification of images is considered to be a relatively simple task for humans, however, the interpretation and classification of these tasks by a machine is considered very complex. The increase in computer processing power, especially GPUs, over the past 20 years has led to significant improvements to both image capture devices and the efficiency of image processing. This has been coupled with the development of newer image processing and computer vision algorithms, which include, among others, the incorporation of machine and deep learning.

Leveraging these advances, a myriad of real-time and highly accurate applications have been developed, especially in the field of medical image processing. In line with this, the improvement in the capturing and analysis of digital images has allowed for images of skin lesions to be digitally captured, analysed and classified using computing technology. There is no generic machine learning algorithm for the multitude of types of source data. Consequently, several classifiers should be tested to empirically validate the given data [8]. The optimal model should be based on the probability of achieving a high degree of success from generalised unseen or test data and not necessarily on being the best classifier for the known data (overfitting).

The methodology and results of similar studies is discussed in section 2. The classification methods and the process of evaluating the classification results are discussed in section 3. Section 4 discusses the selected database and the proposed methodology for classifying the dermoscopic images. The analysis of the classification results is presented in section 5 with concluding remarks in section 6.

II. Related Work

There have been several studies that used the interactive atlas of dermoscopy database to investigate the use of machine learning to classify dermoscopic images. Kasmi and Mokrani [9], used 120 training and 80 test images, employed the ABCD rule for feature extraction and the total dermoscopic score (TDS) for classification and produced an accuracy of 94%, specificity of 96% and sensitivity of 91%, with a false negative rate of 9%. TDS is a simple to implement multi-criteria decision analysis tool. Tschandl et. al [10], using 888 training images and 176 images for testing, employed a convoluted neural network (CNN) using Resnet-50 and produced a classification accuracy of 76%, specificity of 70% and sensitivity of 86%, with a false negative rate of 14%. CNN does not require significant pre-processing but requires a large amount of training data to be able to produce accurate results.

Barata et. al [11] utilised colour normalisation methods, bag of features and K-means clustering for feature extraction on the dermoscopic images. Thereafter, classification using SVM, on 340 test images and produced an accuracy of 75%, specificity of 70% and sensitivity of 81%, yielding a false negative rate of 19%. Only enhanced colour features were

investigated in their study and excluded asymmetry, border and diameter features. This method however improves the colour constancy from multiple image sources.

III. Classification Methods

One common classifier is the k-nearest neighbour (KNN). This is a simple memory-based classification model that calculates the Euclidian distance from a known data point [12]. The classification is determined by the majority of samples in the k neighbours. The value of k determines the number of neighbours used to classify an unknown data point. It is usually chosen to be an odd integer to avoid ties between classes. Equation (1) shows the formula used to calculate Euclidian distance, d_{st}, between two vectors, x_s and y_t [13].

$$d_{st}^2 = (x_s - y_t)(x_s - y_t)' \qquad (1)$$

Another common classifier is the Naïve Bayes which is a probabilistic classifier that works well with a high number of input features [14]. Equation (2) expresses mathematically how Naïve Bayes determines the probability of an event occurring given the probability of a prior event.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \qquad (2)$$

where $P(A|B)$ is the probability of A given B, $P(B|A)$ is the probability of B given A, $P(A)$ is the probability of A, and $P(B)$ is the probability of B.

A third classification methodology is the support vector machine (SVM) which is a common choice for image classification in the biomedical field owing to its ease of use and flexibility [15]. SVMs, which was originally developed as a binary classifier, separates data into classes using linear or non-linear classification depending on the type of data provided. The line or boundary that is used to separate the data into classes is referred to as the hyperplane[16, 17]. The optimal hyperplane maximises a margin between the support vectors and is illustrated in Fig. 2.
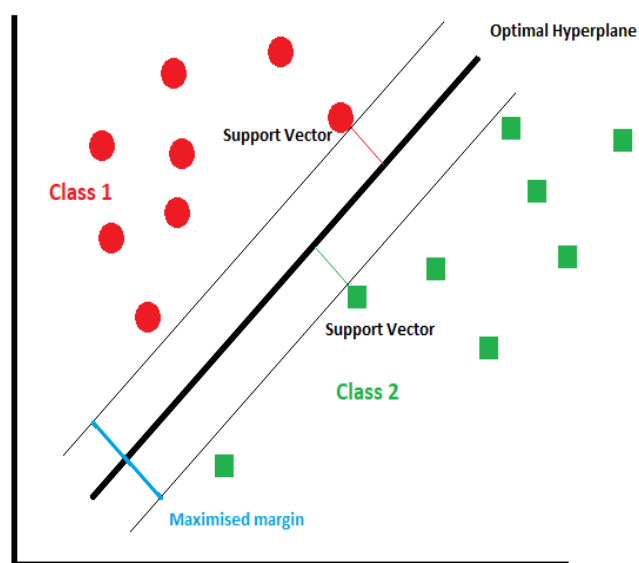

Fig. 2. Support vector machine classification for linearly separable data

The optimal hyperplane is found by maximising the margin (w). This is determined by the perpendicular distance between the support vectors and the hyperplane. Equation (3) is used to calculate a Class 1 support vector.:

$$w.x + b = -1 , \qquad (3)$$

and the Class 2 support vector is given by (4):

$$w.x + b = +1. \qquad (4)$$

The equation for the hyperplane is given by (5):

$$w.x + b = 0 , \qquad (5)$$

where w is the weight vector, x is the input feature vector, and b denotes the bias. Obtaining the largest possible margin will provide a better classification of the data and therefore reduce classification errors.

The evaluation of the classification algorithms is measured by calculating the accuracy, sensitivity and specificity [18]. Accuracy refers to the number of *correct* predictions over the *total* predictions made which is expressed by (6):

$$Accuracy = \frac{t_p + t_n}{t_p + f_p + t_n + f_n} , \qquad (6)$$

where t_p denotes a true positive, t_n refers to a true negative, f_p denotes a false positive, and f_n refers to a false negative.

Sensitivity/Recall (true positive rate) refers to the proportion of actual *positives* that are correctly classified. This is expressed by (7):

$$Sensitivity = \frac{t_p}{t_p + f_n} . \qquad (7)$$

Specificity (true negative rate) is the proportion of actual *negative* samples that are correctly classified and is given as (8):

$$Specificity = \frac{t_n}{t_n + f_p} . \qquad (8)$$

IV. PROPOSED METHODOLOGY

Owing to the nature of the condition under study, the minimisation of false negatives must take precedence over false positives. Depending on the weighting of the machine learning algorithm in the final analysis by the physician, a false negative could be detrimental to a patient should treatment be delayed. On the other hand, a false positive may only be considered as an inconvenience to the patient. Nevertheless, the classification should also ensure false positives are limited as this would result in an unnecessary specialist referral, unnecessary biopsies and undue patient anxiety.

The dermoscopic image dataset (Interactive Atlas of Dermoscopy) used in this research is provided by Simon Fraser University [19]. The images were captured using a Dermaphot/Optotechnik dermoscope at a dimension of 768 x 512 pixels. The dataset contains 160 images of melanoma and non-melanoma cases with 120 images used for training and 40 images used for testing. For this study, five-fold cross-validation was employed for the generation of the independent datasets. The 40 test images were divided evenly for melanoma and non-melanoma cases. This research investigates the early detection of melanoma. Therefore, the images selected from the dataset need to be in stage 0 or stage 1. The melanoma lesions chosen are in-situ or less than 0,76mm in thickness. This means the melanoma is still located on the epidermis and has not spread to deeper skin layers or to other tissue [20].

The images selected require to be complete as the entire border of the lesion is compulsory to allow for segmentation (delineation of the region of interest). An incomplete lesion would affect the asymmetry, border and diameter values in the feature extraction stage as shown in Fig. 3. Images that are out of focus will negatively impact the edge detection for image segmentation.
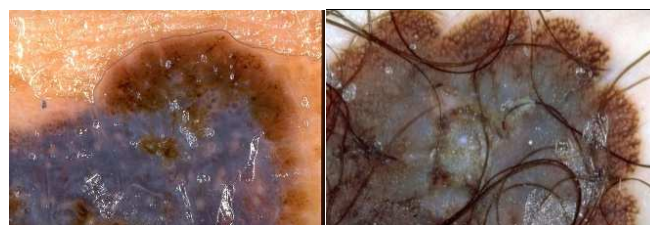
Fig. 3. Sample of images rejected due to being incomplete or out of focus [19]

The images undergo pre-processing, segmentation and feature extraction as depicted in Fig. 4. The pre-processing stage employs a median filter to remove the noise from the images. Histogram stretching is applied to improve contrast. Image segmentation is achieved using morphological processes of opening and closing.

Once the image is segmented, the features are then extracted. The extracted features include the asymmetry index, border index, colour and diameter (ABCD). Fig. 5 shows the graphical user interface used to process the original dermoscopic images to extract the ABCD features needed create the training and test datasets.

Fig. 4. The five stages towards computer aided diagnosis

Fig. 5. GUI used to extract features to create the training and test datasets
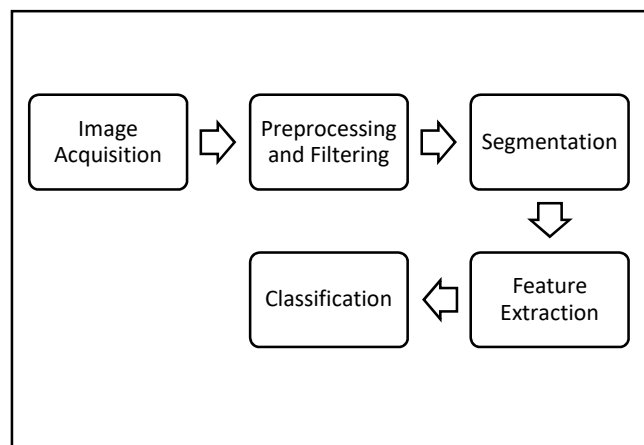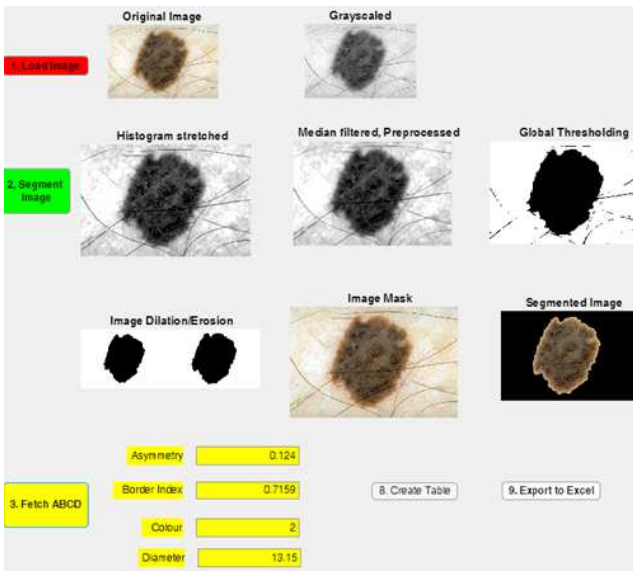
The asymmetry index is calculated by finding the centroid of the segmented image and folding the image along the x and y axis and calculates the difference in area. The border irregularity is found by computing the isoperimetric quotient. Colour was determined by counting the number of colours of interest that are present in the lesion. The diameter is found by the two pixels in the lesion that are the furthest away from each other and passing the centroid, namely the major axis.

The *k*-nearest neighbours (KNN), Naïve Bayes, and a linear support vector machine (LSVM) models are trained using MATLAB® and subsequently tested using the generated prediction scripts. The KNN classifier used an initial value of K=10 and the Euclidian distance metrics to calculate the distance between the test and training sample. The distances are ranked from nearest to furthest and the top 10 are used. Tied values, equal to the 10th value, are also included. The value of K was incremented from 1-15 and a value of 10 provided the lowest false negative rate of the test samples. The test samples are classified based on the majority class of the 10 nearest neighbours. If a tie occurs between the 10 nearest neighbours, the class with the nearest neighbour is used.

Naïve Bayes was implemented using a Gaussian (normal) distribution. The mean and standard deviation of the training data is calculated by estimating the normal distribution for each class. The test data classification is predicted by calculating the posterior probability of a particular test sample belonging to each class. The test sample is then allocated to a class based on the highest posterior probability.

The data is not perfectly separable, therefore LSVM uses a soft margin with the box constraint or penalty factor set to 1 to allow for miscalculations and to avoid overfitting. The kernel scale is set to 2 to reduce bias towards features with higher magnitudes. Classification performance of each algorithm is measured by calculating the accuracy, sensitivity, specificity and generating a confusion matrix [21]. This is discussed in the next section.

V. RESULTS

The four extracted features, viz. asymmetry, border, colour and diameter, are used to generate the training and testing datasets. The training datasets in turn are used to create the KNN, Naïve Bayes, and LSVM models. Finally, the test datasets are inputted into the trained models for classification.

Fig. 6. Confusion Matrix for k-Nearest Neighbours for 40 test samples

Fig. 7. Confusion Matrix for Naïve Bayes for 40 test samples

Fig. 8. Confusion Matrix for Linear SVM for 40 test samples

Fig. 6 shows the decision matrix for the KNN model. In this case, the model correctly classified 30 out of the 40 images providing an accuracy of 75% and produced 4 false negative classifications. Fig. 7 illustrates the decision matrix for the Naïve Bayes model. The model correctly classified 29 out of 40 images providing an accuracy of 72.5% and also produced 4 false negative classifications. Fig. 8 shows the decision matrix for the Linear SVM. In this case, the model correctly classified 34 out of 40 images providing an accuracy of 85% and produced only 1 false negative classification. Table 1 provides a statistical breakdown of the results. The true positive rate (sensitivity) of 95% obtained for Linear SVM outperforms the 81% achieved for both Naïve Bayes and KNN.

TABLE I.   STATISTICAL MEASUREMENT OF RESULTS

	Naïve Bayes	KNN	Linear SVM
Sensitivity	81%	81%	95%
Specificity	63%	68%	74%
Accuracy	73%	75%	85%
False Positive Rate	37%	32%	26%
False Negative Rate	19%	19%	5%

Moreover, the true negative rate (specificity) of 74% obtained for Linear SVM was notably greater than the 63% obtained for Naïve Bayes and the 68% obtained for KNN. Sensitivity is an indicator of how well the algorithms can correctly detect melanoma. A high sensitivity shows that the algorithm has a lower chance of providing a false negative result and is ideal for positively diagnosing a condition, namely melanoma. KNN and Naïve Bayes produced a false negative rate of 19% and Linear SVM produced a comparatively low 5% false negative rate. Therefore, a non-melanoma classification result would be reliable. A high specificity value will produce a low false-positive rate and is an indicator of the reliability of the melanoma classification result [8].

VI. CONCLUSION

This research presents the 5 stages towards computer aided diagnosis of skin lesions using dermoscopic images, namely, image acquisition, pre-processing, segmentation, feature extraction and finally classification. The dermoscopic images chosen are of melanoma lesions less than 0,76mm in thickness which corresponds to the early stages of cancer. The ABCD rule, Asymmetry, Border, Colour and Diameter was used as the guideline to provide the features for extraction. The images underwent binary classification of skin lesions as either melanoma or non-melanoma through the use of machine learning. Three classifiers were chosen which include; k-nearest neighbours (KNN), Naïve Bayes, and a linear support vector machine (LSVM). The performance of the selected classification algorithms is compared based on accuracy, sensitivity and specificity. Early detection of melanoma is vital to improving the chances of survival. Therefore, minimising false negative classification is very important to ensuring the patient receives the necessary treatment without delay. All classifiers used in this study have been able to differentiate the images into their respective class with varying degrees of success.

The experimental results show:

- Linear SVM algorithm was the optimal classifier with an accuracy of 85%, specificity of 74% and sensitivity of 95%. The false negative rate was 5%.
- Naïve Bayes algorithm produced an accuracy of 73%, specificity of 63% and sensitivity of 81%. The false negative rate was 19%.
- KNN algorithm produced an accuracy of 75%, specificity of 68% and sensitivity of 81%. Similar to Naïve Bayes, the false negative rate was 19%.

The classification of the lesions using classification algorithms may be considered as part of the decision methods employed by dermatologist and physicians. This has the potential to reduce the number of false negative cases and potentially aid in the earlier diagnosis and treatment of melanoma.

The shortcomings of this research are largely due to the quality and size of the available dataset. Several images were excluded from the study due to poor capturing techniques which reduced the size of the dataset. A larger, high quality dataset would enable the use of neural networks which require large training data.

Future work will include training and testing the models on a significantly larger dataset as well as include hybrid methods using the ABCD rule in conjunction with other rules such as Menzies and a seven-point checklist. Furthermore, the inclusion of metadata, such as age, location of lesion, and evolution of lesion, together with the datasets, may be considered for the improvement in the accuracy of the classifiers.

VII. REFERENCES

[1] C. Y. Wright, M. Norval, B. Summers, L. M. Davids, G. Coetzee, and M. Oriowo, "Solar ultraviolet radiation exposure and human health in South Africa: finding a balance," *SAMJ: South African Medical Journal,* vol. 102, 2012, pp. 665-666.

[2] C. Conforti, R. Giuffrida, R. Vezzoni, F. S. S. Resende, N. di Meo, and I. Zalaudek, "Dermoscopy and the experienced clinicians," *International Journal of Dermatology,* vol. 59, no. 1, 2020, pp. 16-22.

[3] R. J. Friedman, D. S. Rigel, and A. W. Kopf, "Early detection of malignant melanoma: The role of physician examination and self-examination of the skin," *CA: A Cancer Journal for Clinicians,* vol. 35, no. 3, 1985, pp. 130-151.

[4]A. C. Halpern. "Melanoma Warning Signs." https://www.skincancer.org/skin-cancer-information/melanoma/melanoma-warning-signs-and-images/ (accessed 4 February 2021).

[5]D. Whitaker and W. Sinclair, *Guideline on the Management of Melanoma* (2008, no. 8). SA Medical Association Health and Medical Publishing Group, , 2008.

[6]M. E. Celebi, N. Codella, A. Halpern, and D. G. Shen, "Guest Editorial Skin Lesion Image Analysis for Melanoma Detection," *Ieee J Biomed Health,* vol. 23, no. 2, 2019, pp. 479-480.

[7]P. Carli, V. De Giorgi, P. Nardini, F. Mannone, D. Palli, and B. Giannotti, "Melanoma detection rate and concordance between self-skin examination and clinical evaluation in patients attending a pigmented lesion clinic in Italy," *Br J Dermatol,* vol. 146, no. 2, pp. 261, 2002.

[8]F. Ian, G. Rayid, S. J. Ron, K. Frauke, and L. Julia, *Big Data and Social Science : A Practical Guide to Methods and Tools*, 2017.

[9]R. Kasmi and K. Mokrani, "Classification of malignant melanoma and benign skin lesions: Implementation of automatic ABCD rule," *IET Image Processing,* vol. 10, 2016.

[10]P. Tschandl, G. Argenziano, M. Razmara, and J. Yap, "Diagnostic accuracy of content-based dermatoscopic image retrieval with deep classification features," *British Journal of Dermatology,* vol. 181, no. 1, 2019, pp. 155-165.

[11]C. Barata, M. E. Celebi, and J. Marques, "Toward a Robust Analysis of Dermoscopy Images Acquired under Different Conditions", 2015, pp. 1-22.

[12]R. Troncy, B. Huet, and S. Schenk, *Machine Learning Techniques for Multimedia Analysis*. 2011, pp. 59-80.

[13]K. Chomboon, P. Chujai, P. Teerarassammee, K. Kerdprasop, and N. Kerdprasop, *An Empirical Study of Distance Metrics for k-Nearest Neighbor Algorithm*. 2015, pp. 280-285.

[14]C. D. Manning, P. Raghavan, and H. Schütze, Introduction to information retrieval / Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze. 2008.

[15]D. A. Pisner and D. M. Schnyer, "Chapter 6 - Support vector machine," in *Machine Learning*, A. Mechelli and S. Vieira Eds.: Academic Press, 2020, pp. 101-121.

[16]M. Somvanshi, P. Chavan, S. Tambade, and S. V. Shinde, "A review of machine learning techniques using decision tree and support vector machine," in *2016 International Conference on Computing Communication Control and automation (ICCUBEA)*, 2016, pp. 1-7.

[17]J. Bell, Machine Learning: Hands-On for Developers and Technical Professionals. John Wiley & Sons Inc, 2014, p. 408.

[18]A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics,* vol. 17, no. 1, 2021, pp. 168-192.

[19]J. Kawahara, S. Daneshvar, G. Argenziano, and G. Hamarneh, "Seven-Point Checklist and Skin Lesion Classification Using Multitask Multimodal Neural Nets," (in English), *Ieee J Biomed Health,* vol. 23, no. 2, 2019, pp. 538-546.

[20]M. Pérez-Ortiz, A. Sáez, J. Sánchez-Monedero, P. A. Gutiérrez, and C. Hervás-Martínez, "Tackling the ordinal and imbalance nature of a melanoma image classification problem," in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 2156-2163.

[21]J. Karakaya, "Evaluation of binary diagnostic tests accuracy for medical researches," *Turkish Journal of Biochemistry / Turk Biyokimya Dergisi,* vol. 46, no. 2, 2021, pp. 103-113.