

Intrinsic Melanin and Hemoglobin Colour Components for Skin Lesion Malignancy Detection

Ali Madooei, Mark S. Drew, Maryam Sadeghi, and M. Stella Atkins

School of Computing Science
Simon Fraser University
amadooei@cs.sfu.ca
<http://www.cs.sfu.ca/~amadooei>

Abstract. In this paper we propose a new log-chromaticity 2-D colour space, an extension of previous approaches, which succeeds in removing confounding factors from dermoscopic images: (i) the effects of the particular camera characteristics for the camera system used in forming RGB images; (ii) the colour of the light used in the dermoscope; (iii) shading induced by imaging non-flat skin surfaces; (iv) and light intensity, removing the effect of light-intensity falloff toward the edges of the dermoscopic image. In the context of a blind source separation of the underlying colour, we arrive at intrinsic melanin and hemoglobin images, whose properties are then used in supervised learning to achieve excellent malignant vs. benign skin lesion classification. In addition, we propose using the geometric-mean of colour for skin lesion segmentation based on simple grey-level thresholding, with results outperforming the state of the art.

1 Introduction

The three most common malignant skin cancers are basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and melanoma, among which melanoma is the most deadly with a high increasing rate in most parts of the world. Melanoma is often treatable if detected in the early stage, particularly before the metastasis phase. Therefore, there is an increasing demand for computer-aided diagnostic systems to catch early melanomas.

Colour has played a crucial role in the diagnosis of skin lesions by experts in most clinical methods (see e.g. [1]). For instance, the presence of multiple colours with an irregular distribution can signal malignancy.

Few studies have investigated the use of colour features representing biological properties of skin lesions. In particular, the work of Claridge et al. has figured prominently, with emphasis on the use of intermediate multispectral modelling to generate images disambiguating dermal and epidermal melanin, thickness of collagen, and blood [2]. At the same time, another stream of work has focused on using Independent Component Analysis (ICA) [3] in the context of 3-channel RGB images with no intermediate spectral-space model, aimed both at non-medical images and dermoscopic images of skin [4].

Here we concentrate on the latter, simpler, approach to utilizing colour and consider only RGB, not multispectral image modelling. We show that, combined with texture features, one can successfully carry out classification, disambiguating Malignant vs. Benign; Melanoma vs. Benign; and Melanoma vs. Spitz Nevus.

2 Method

We first adopt the ICA-based idea [4] in spirit and show that, in a particular novel colour space, pixel triples live on a plane, with (non-orthogonal) basis vectors assumed attributable to melanin and hemoglobin only.

Here, in an innovative step, we introduce a new colour 2-D chromaticity which removes (i) the effects of the particular camera characteristics for the camera system used in forming RGB images; (ii) the colour of the light used in the dermoscope; (iii) shading induced by imaging non-flat skin surfaces; (iv) and also light intensity, removing the effect of light-intensity falloff toward the edges of the image. The output from this colour processing is a set of two 1-D-colour chromaticity images, one for melanin content and one for hemoglobin content.

Together with the above colour space features, we also employ greyscale and texture features, including all features in a final 25-D feature-space vector. Such vectors are then amenable to machine learning techniques for effective skin lesion classification. In this paper we achieve comparable to state of the art results for distinguishing malignant from benign lesions.

2.1 Colour Space Image Formulation

Tsumura et al. first suggested using a simple Lambert-Beer type of law for radiance from a multilayer skin surface, resulting from illumination by polarized light [5]. That is, employing a model similar to a simple logarithm model based on optical densities for accounting for light passing for example through multilayer slide film. The transmittance through each colour layer is proportional to the exponential of the negative optical density for that layer. Such a simple model stands in contradistinction to a considerably more complex model based on Kubelka-Monk (KM) theory such as used in [2]. In the latter, full modelling of interreflection inside each layer is used to set up equations detailing light transport. This uses estimates of the absorption K and scattering S in each layer to predict overall transmittance and reflection [6]. KM theory has been found to be useful in tasks such as visualizing different components including surface and deep melanin etc. [2]. Here, we are simply focused on the classification task, and make use of the simpler model.

In the simpler approach, then, we utilize the model developed by Hiraoka et al. [7], which formulates a generalization of the Lambert-Beer law. In [7], the spectral reflection of skin (under polarized light) at pixel indexed by (x, y) is given by

$$S(x, y, \lambda) = \exp\{-\rho_m(x, y)\alpha_m(\lambda)l_m(\lambda) - \rho_h(x, y)\alpha_h(\lambda)l_h(\lambda) - \zeta(\lambda)\} \quad (1)$$

where $\rho_{m,h}$ are densities of melanin and hemoglobin respectively (cm^{-3}), and are assumed to be independent of each other. The cross sectional areas for scattering absorption of melanin and hemoglobin are denoted $\alpha_{m,h}$ (cm^2) and $l_{m,h}$ are the mean pathlength for photons in epidermis and dermis layers, which are used as the depth of the medium in this modified Lambert-Beer law. These quantities are used as well in [4]. Finally, we also extend the model by including a term ζ standing for scattering loss and any other factors which contribute to skin appearance such as absorbency of other chromophores (e.g. β -carotene) and thickness of the subcutis. The reason we can extend the model will become clear below, when we form logarithms of ratios in a novel step.

In keeping with [8] we adopt a standard model in computer vision for colour image formation. Suppose the illuminant spectral power distribution is $E(\lambda)$ and, in any reflective case, the spectral reflectance function at pixel (x, y) is $S(x, y, \lambda)$, e.g. as given in eq. (1) above. Then measured RGB values are given by

$$R_k(x, y) = \omega(x, y) \int E(x, y, \lambda_k) S(x, y, \lambda_k) Q_k(\lambda) d\lambda, \quad k = 1..3 \quad (2)$$

where ω denotes shading variation (e.g., Lambertian shading is surface normal dotted into light direction, although we do not assume Lambertian surfaces here); and $Q_k(\lambda)$ is the camera sensor sensitivity functions in the R,G,B channels.

Following [8] we adopt a simple model for the illuminant: we assume the light can be written as a Planckian radiator (in Wien's approximation):

$$E(x, y, \lambda, T) \simeq I(x, y) k_1 \lambda^{-5} \exp(-k_2/(T\lambda)) \quad (3)$$

where k_1 and k_2 are constants, T is the correlated colour temperature characterizing the light spectrum, and I is the lighting intensity at pixel (x, y) , allowing for a possible rolloff in intensity towards the periphery of the dermoscopic image. We assume light temperature T is constant across the image (but is, in general, unknown).

Finally, with [8] we assume camera sensors are narrowband or can be made narrowband via a spectral sharpening operation [9]. In this approximation, sensor curve $Q_k(\lambda)$ is simply assumed to be a delta function: $Q_k(\lambda) = q_k \delta(\lambda - \lambda_k)$, where specific wavelengths λ_k and sensor-curve heights q_k are properties of the camera used. Simplifying by taking logs (cf. [4]), we arrive at a model for pixel log-RGB as follows:

$$\begin{aligned} \log R_k(x, y) = & -\rho_m(x, y) \sigma_m(\lambda_k) - \rho_h(x, y) \sigma_h(\lambda_k) - \zeta(\lambda_k) \\ & + \log(k_1 I(x, y) \omega(x, y)) + [\log(1/\lambda_k^5) - k_2/(\lambda_k T)] \end{aligned} \quad (4)$$

where we have lumped terms $\sigma_m(\lambda_k) = \alpha_m(\lambda_k) l_m(\lambda_k)$, $\sigma_h(\lambda_k) = \alpha_h(\lambda_k) l_h(\lambda_k)$. For notational convenience, denote $u_k = \log(1/\lambda_k^5)$, $e_k = -k_2/\lambda_k$, $m_k = \sigma_m(\lambda_k)$, $h_k = \sigma_h(\lambda_k)$, $\zeta_k = \zeta(\lambda_k)$.

Now let us move forward from [4] by making the novel observation that the same type of chromaticity analysis as appears in [8] can be brought to bear here for the skin-reflectance model (4) [but N.B., [8] does not use the density model (1)]. Chromaticity is colour without intensity, e.g. an L_1 -norm based chromaticity is $\{r, g, b\} = \{R, G, B\}/(R + G + B)$. Here, suppose we instead form a band-ratio chromaticity by dividing by one colour-channel R_p , e.g. Green for $p = 2$. [In practice, we shall instead follow [8] and divide by the geometric-mean colour, $\mu = \sqrt[3]{R \cdot G \cdot B}$, so as not to favour one particular colour-channel, but dividing by R_p is clearer in exposition.] Notice that dividing removes the effect of shading ω and light-intensity field I .

Defining a log-chromaticity $\chi(x, y)$ as the log of the ratio of colour component R_k over R_p , we then have

$$\begin{aligned} \chi_k(x, y) = & \log(R_k(x, y)/R_p(x, y)) \\ = & -\rho_m(x, y)(m_k - m_p) - \rho_h(x, y)(h_k - h_p) + w_k - (e_k - e_p)(1/T) \end{aligned} \quad (5)$$

with $w_k \equiv (u_k - u_p) - (\zeta_k - \zeta_p)$. The meaning of this equation is that, if we were to vary the lighting (in this simplified model) then the chromaticity χ would follow a

straight line as temperature T changes. In fact, this linear behaviour is also obeyed by the mean $\bar{\chi}$ over the image of this new chromaticity quantity:

$$\bar{\chi}_k = -\bar{\rho}_m(m_k - m_p) - \bar{\rho}_h(h_k - h_p) + w_k - (e_k - e_p)(1/T) \quad (6)$$

Now we notice that we can remove all terms in the camera-offset term w_k and the illuminant-colour term T by subtracting the mean from χ . Let χ^0 be the mean-subtracted vector $\chi_k^0(x, y) = \chi_k(x, y) - \bar{\chi}_k$. We then arrive at a feature which depends only on melanin m and hemoglobin h :

$$\chi_k^0(x, y) = -(\rho_m(x, y) - \bar{\rho}_m)(m_k - m_p) - (\rho_h(x, y) - \bar{\rho}_h)(h_k - h_p) \quad (7)$$

If we apply the assumption that m and h terms can be disambiguated using ICA, then from the new feature χ^0 we can extract the melanin and hemoglobin content in dermoscopic images, where we take vectors $(m_k - m_p)$ and $(h_k - h_p)$ as constant vectors in each image. The log-subtraction step removes intensity and shading, and the mean-subtraction removes camera-offset and light colour, as opposed to [4] where one must attempt to recover approximations of these quantities.

As an example, consider Fig. 1(a) showing a Melanoma lesion, and the ρ_m and ρ_h components in Figs.(b,c). Below, we show how these two new image features, $\rho_m^0(x, y) = (\rho_m(x, y) - \bar{\rho}_m)$ and $\rho_h^0(x, y) = (\rho_h(x, y) - \bar{\rho}_h)$, can be used in lesion classification. In computer vision, images with lighting removed are denoted “intrinsic images”, and thus our two new features are indeed intrinsic.

Geometric Mean Chromaticity. To not rely on any particular colour channel, we divide not by R_p but by the geometric mean μ at each pixel, for which the invariance properties above persist: $\psi_k(x, y) \equiv \log[R_k(x, y)/\mu(x, y)]$. Then ψ is a 3-vector; it is orthogonal to $(1, 1, 1)$. Therefore instead of 3-vectors one can easily treat these as 2-vector values, lying in the plane orthogonal to $(1, 1, 1)$: if the 3×3 projector onto that 2-D subspace is P , then the singular value decomposition of $P = UU^T$, where U is a 3×2 matrix. We project onto 2-D vectors ϕ in the plane coordinate system via U^T :

$$\psi_k(x, y) = \log[R_k(x, y)/\mu(x, y)]; \quad \phi = U^T \psi \quad (8)$$

where ϕ is 2-D. The mean-subtraction above still holds in projected colour, and we therefore here propose carrying out ICA in the plane: $feature = \eta = ICA(\phi - \bar{\phi})$.

2.2 Texture and Colour Feature Vectors

So far, we have discarded the luminance (intensity) part of the input image, focusing on intrinsic colour. However, we can go on to include the greyscale geometric-mean image (Fig. 1(d)) information μ as well. Thus, we extract features for each of $\{\mu, \eta_1, \eta_2\}$.

As colour features, we generate mean; standard deviation; the ratio of these; and entropy of each channel, in addition to $|var(\eta_1) - var(\eta_2)|$, adding up to a 13-D colour feature vector. Further, we add texture features to our colour feature-vectors, in a similar fashion as in [10]: four of the classical statistical texture measures of [11] (contrast, correlation, homogeneity and energy) are derived from the grey level co-occurrence matrix (GLCM) of each channel. This is an additional 12-D texture feature vector; thus we arrive at a 25-D feature vector.

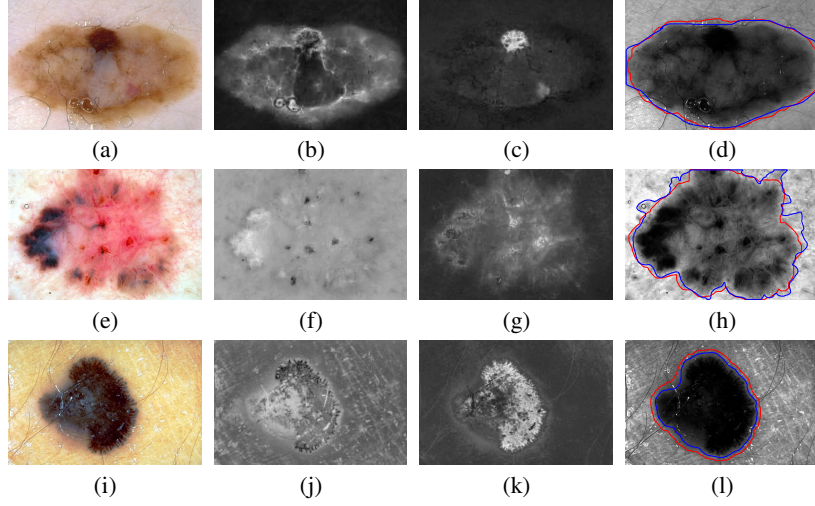


Fig. 1. (a): Input image (Melanoma); (b): melanin component; (c): hemoglobin component; (d): greyscale geometric mean. Blue border: expert segmentation, Red border: our segmentation. (e-h): BCC. (i-l): Spitz Nevus.

3 Image Masks

Each of the features calculated above is applied only within a mask surrounding the lesion, normalized accordingly. For automatic segmentation of lesions, we found that using the geometric-mean μ is as good as or better than the state of the art [12] for these dermoscopic images, in a much simpler algorithm. Here we simply apply Otsu's method [13] for selecting a grey-level threshold. Note that Otsu's method (and also most commercially available automated systems) fail in segmenting low contrast lesions. However our approach achieved very high precision and recall, since we discovered that geometric-mean greyscale highlights the lesion from its surrounding.

We tested our method on a dataset of images used by Wighton et al. [12]. They presented a modified random walker (MRW) segmentation where seed points were set automatically based on a lesion probability map (LPM). The LPM was created through a supervised learning procedure using colour/texture properties. Table 1 shows results for our method compared to results in [12]. While our method for segmentation uses a much simpler algorithm and does not require learning, it achieves competitive results. It is worth mentioning [12] also applied Otsu's method on their lesion probability maps. Their result included in Table 1 under 'Otsu on LPM', with results not nearly as good as ours. In another test on **944 test images**, we achieved precision 0.86, recall 0.95, and f-measure 0.88 (with STD 0.19, 0.08 and 0.15 respectively) compared to expert segmentations.

4 Experiments

We applied a Logistic classifier to a set of 500 images, with two classes consisting of malignant (melanoma and BCC) vs. all benign lesions (congenital, compound, dermal,

Table 1. Comparing our segmentation method to the modified random walker (MRW) algorithm and Otsu's thresholding, on lesion probability map (LPM)[12]. The dataset is divided into a set of 20 easy-to-segment images, and another 100 images that pose a challenge to segmentation methods. Note that our method consistently produces higher f-measures.

ImageSet	n	Method	Precision	Recall	F-measure
simple	20	MRW on LPM	0.96	0.95	0.95
		Otsu on LPM	0.99	0.86	0.91
		Our Method (STD)	0.94 (0.04)	0.97 (0.04)	0.95 (0.02)
challenging	100	MRW on LPM	0.83	0.90	0.85
		Otsu on LPM	0.88	0.68	0.71
		Our Method (STD)	0.88 (0.15)	0.90 (0.1)	0.88 (0.09)
whole	120	MRW on LPM	0.87	0.92	0.88
		Otsu on LPM	0.91	0.74	0.78
		Our Method (STD)	0.89 (0.13)	0.90 (0.09)	0.89 (0.09)

Table 2. Results of classifying the dataset using different colour spaces. MHG is our proposed colour space; We win and improve the f-measure somewhat, but the AUC is substantially boosted. Since our dataset is unbalanced, a classifier trained on e.g. RGB achieved high score while assigned benign label to most malignant instances. We on the other hand produced equally high and steady results for both classes; improving e.g. recall for malignant cases up to 23%. Since same feature-set & classifier is used, the improvement is the result of using our proposed colour-space.

Colour Space	Class	n	Precision	Recall	F-measure	AUC
MHG	Malignant	135	0.806	0.8	0.803	0.953
	Benign	365	0.926	0.929	0.927	
	Weighted Avr.	500	0.894	0.894	0.894	
RGB	Malignant	135	0.895	0.57	0.697	0.773
	Benign	365	0.86	0.975	0.914	
	Weighted Avr.	500	0.869	0.866	0.855	
HSV	Malignant	135	0.807	0.652	0.721	0.797
	Benign	365	0.88	0.942	0.91	
	Weighted Avr.	500	0.86	0.864	0.859	
LAB	Malignant	135	0.837	0.57	0.678	0.765
	Benign	365	0.858	0.959	0.906	
	Weighted Avr.	500	0.852	0.854	0.844	

Clark, Spitz and blue nevus; dermatofibroma; and seborrheic keratosis). Table 2 results are averaged over 10-fold cross-validation. We achieve **f-measure: 89.4%** and **AUC: 0.953**, an excellent performance. For comparison, we compare using our feature set on RGB, HSV, and CIELAB colour spaces. We see that our proposed colour space, $\{\eta_1, \eta_2, \mu\}$ (denoted MHG for melanin, hemoglobin and geometric-mean), improves accuracy (f-measure) as well as the performance (AUC) of classification, particularly formative for malignant lesions, where the results show significantly higher precision and recall for our method.

Table 3. Results of classifying the dataset using different subsets of our feature-set (colour/texture), and different channels of our proposed colour space MHG

Description	n	Precision	Recall	F-measure	AUC
Colour Features Only on MHG	500	0.728	0.754	0.72	0.731
Texture Features only on MHG		0.855	0.854	0.842	0.858
Colour+Texture on Melanin only		0.783	0.794	0.786	0.831
Colour+Texture on Hemoglobin only		0.765	0.78	0.766	0.829
Colour+Texture on Geo-mean only		0.817	0.824	0.817	0.877
Colour+Texture on MHG		0.894	0.894	0.894	0.953

Table 4. Classification results for skin cancer categories using our proposed feature space

Classification Task	n	Precision	Recall	F-measure	AUC
Malignant vs. Benign	500	0.894	0.894	0.894	0.953
Melanoma vs. Benign	486	0.897	0.899	0.897	0.946
Melanoma vs. Spitz Nevus	167	0.915	0.916	0.916	0.96

To judge the effect of colour vs. texture and the different channels of our proposed colour space MHG, Table 3 shows that 1) texture features have higher impact than colour features; 2) the three channels of MHG contribute more than each individually; best overall is from combining all.

To further analyze the robustness and effectiveness of our method, we tried different classifiers using Weka [14]. On the main classification task of malignant vs. benign, Logistic Regression produced the highest result (Table 4) whereas e.g. using support vector machine (SVM) we attained precision 0.872, recall 0.87, f-measure 0.871 and AUC 0.828; sequential minimal optimization (SMO) produced 0.892, 0.891, 0.888, 0.883 respectively. Table 4 shows results for classifying melanoma vs. benign and melanoma vs. Spitz nevus, as well as malignant vs. benign, with excellent results for these difficult problems. Spitz nevus is a challenging classification, to the extent that expert dermatologists usually have to take into consideration other criteria such as patient's age.

5 Conclusion

We have proposed a new colour-feature η which is aimed at apprehending underlying melanin and hemoglobin biological components of dermoscopy images of skin lesions. The advantage of the new feature, in addition to its biological underpinnings, lies in removing the effects of confounding factors such as light colour; intensity falloff; shading; and camera characteristics. The new colour-feature vectors $\{\eta_1, \eta_2\}$ combined with geometric-mean vector, μ , is proposed as a new colour-space MHG (abbreviation of melanin, hemoglobin and geometric-mean). In our experiments, MHG is shown to produce excellent results for classification of Malignant vs. Benign; Melanoma vs. Benign; and Melanoma vs. Spitz Nevus. Moreover, in the lesion segmentation task, μ is shown to improve accuracy of segmentation. Future work will include i) Exploration of effects and contributions of other colour and texture features, combined with those reported here. ii) Experimenting with different learning algorithms and strategies, in particular the possibility of multi-class classification. iii) Examination of the extracted melanin

and hemoglobin colour components as a set of two full-colour images, since the equations leading to (8) are in fact invertible for each component separately. As 3-D colour features these will support descriptors such as colour histograms and correlograms, which may lead to even more improvement.

References

1. Henning, J.S., Dusza, S.W., Wang, S.Q., Marghoob, A.A., Rabinovitz, H.S., Polsky, D., Kopf, A.W.: The CASH (color, architecture, symmetry, and homogeneity) algorithm for dermoscopy. *J. of the Amer. Acad. of Dermatology* 56, 45–52 (2007)
2. Claridge, E., Cotton, S., Hall, P., Moncrieff, M.: From colour to tissue histology: Physics-based interpretation of images of pigmented skin lesions. *Med. Im. Anal.* 7, 489–502 (2003)
3. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley and Sons, Inc., New York (2001)
4. Tsumura, N., Ojima, N., Sato, K., Shiraishi, M., Shimizu, H., Nabeshima, H., Akazaki, S., Hori, K., Miyake, Y.: Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin. *ACM Trans. Graph.* 22, 770–779 (2003)
5. Tsumura, N., Haneishi, H., Miyake, Y.: Independent-component analysis of skin color image. *J. of the Optical Soc. of Amer. A* 16, 2169–2176 (1999)
6. Kang, H.R.: Color technology for electronic imaging systems. SPIE Optical Eng. Press (1997)
7. Hiraoka, M., Firbank, M., Essenpreis, M., Cope, M., Arridge, S.R., Zee, P.V.D., Delpy, D.T.: A Monte Carlo investigation of optical pathlength in inhomogeneous tissue and its application to near-infrared spectroscopy. *Phys. Med. Biol.* 38, 1859–1876 (1993)
8. Finlayson, G.D., Drew, M.S., Lu, C.: Intrinsic Images by Entropy Minimization. In: Pajdla, T., Matas, J.(G.) (eds.) ECCV 2004, Part III. LNCS, vol. 3023, pp. 582–595. Springer, Heidelberg (2004)
9. Finlayson, G.D., Drew, M.S., Funt, B.V.: Spectral sharpening: sensor transformations for improved color constancy. *J. Opt. Soc. Am. A* 11(5), 1553–1563 (1994)
10. Sadeghi, M., Razmara, M., Wighton, P., Lee, T., Atkins, M.: Modeling the Dermoscopic Structure Pigment Network Using a Clinically Inspired Feature Set. In: Liao, H., Eddie Edwards, P.J., Pan, X., Fan, Y., Yang, G.-Z. (eds.) MIAR 2010. LNCS, vol. 6326, pp. 467–474. Springer, Heidelberg (2010)
11. Haralick, R.M., Shapiro, L.G.: Computer and Robot Vision, vol. 1, p. 459. Addison-Wesley, New York (1992)
12. Wighton, P., Sadeghi, M., Lee, T.K., Atkins, M.S.: A Fully Automatic Random Walker Segmentation for Skin Lesions in a Supervised Setting. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 1108–1115. Springer, Heidelberg (2009)
13. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Trans. on Systems, Man and Cybernetics* 9(1), 62–66 (1979)
14. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: WEKA data mining software (2001), <http://www.cs.waikato.ac.nz/ml/weka/>