A Border Irregularity Measure Using Hidden Markov Models as a Malignant Melanoma Predictor

Benjamin S. Aribisala and Ela Claridge *

School of Computer Sciences, The University or Birmingham, Birmingham B15 2TT, U.K.

Abstract. Malignant melanoma, a skin cancer, manifests itself as a dark lesion, most often with an irregular boundary. The degree of irregularity is an important diagnostic indicator. This paper presents a new measure of irregularity using Hidden Markov Models (HMMs) based on the Weibull probability distribution. The measure was tested on 98 skin lesions of which 16 were malignant melanoma. The ROC analysis showed that the measure is 82% sensitive and 82% specific in discriminating the malignant and benign lesions. These results compare favourably with other measures and indicate that HMM captures some distinguishing features in the boundary of malignant lesions.

1 Introduction

Malignant melanoma can be characterised using physical features such as shape, edge, colour and surface texture of skin lesions. The border irregularity of pigmented skin is a significant factor in clinical diagnosis of melanoma [1]. This fact was supported by evidences from medical textbooks [2]. Additionally, border irregularity is a major feature in the seven point checklist used for computing a "suspiciousness" score for skin lesions [3].

It has been empirically discovered that clinicians have difficulties in visually assessing border irregularity of skin lesion outlines and that their assessments are not invariant to reflection and rotation [2, 4]. Much research on quantitative measures of irregularity has been carried out to overcome these shortcomings [2]. The most common approaches include the Compactness Index (e.g. [5]), Fractal Dimension (e.g. [6]) and measures based on radial distance (e.g. [7]). These methods are critically reviewed in a recent paper by Lee et al. [8].

The term "*Irregularity*" is intuitive and can express different meanings. If irregularity is to be quantified, it is necessary first to develop its formal definition, or at least provide its formal description. Five attributes of irregularity have been proposed [9]. One of these attributes which is of interest here is lack of predictability. The elements of a sequence corresponding to a regular shape or pattern are predictable, whereas in an irregular sequence they cannot be easily predicted. That is, the extent to which a sequence can be predicted may help us to determine how regular the sequence is. This paper presents a new measure of irregularity based on Hidden Markov Model. In contrast to the existing measures, the proposed measure is based on a formal criterion of irregularity outlined above.

Section 2 presents a brief description of the HMMs while section 3 presents the HMM for skin lesions. Section 4 describes the experiments. Results and discussion are in section 5. Finally, section 6 presents the conclusion.

2 Hidden Markov Model

A Markov model can be defined as a process that consists of a finite number of states N, and an N x N stochastic matrix with elements a_{ij} which gives the probabilities of transition from states *i* to *j*. A Hidden Markov Model (HMM) is a "composite" Markov model where each state has an associated probability density function $b_j(O_t)$ which gives the probability of the state *j* emitting a particular observation O_t at time t. The states are hidden and can only be observed through the emissions.

Let $\{S_t\}$ and $\{O_t\}$ be sequences of states and associated observations respectively, t=1, ..., T where T is the sequence length. HMM assumes that S_t depends only on S_{t-1} and that the observation O_t is independent of any other observations. Therefore the joint probability distribution of $\{S_t\}$ and $\{O_t\}$ can be written as [10]

$$P(S_1, \cdots, S_T, O_1, \cdots, O_T) = P(S_1)P(O_1/S_1)\prod_{t=2}^{T} P(S_t/S_{t-1})P(O_t/S_t) \quad (1)$$

Using the observation variables HMMs can be discrete-valued or continuous valued. For a continuous HMM which is of interest here $P(O_t/S_t)$ can be modelled using different probability distributions such as Gaussian, Gamma,

* B.S.Aribisala@cs.bham.ac.uk and E.Claridge@cs.bham.ac.uk

exponential, Weibull etc. In this paper we model the observation sequence using Weibull distribution (see 3.2). For a detailed discussion of HMMs see [10].

3 Skin Lesion Shape Model

3.1 Data Description and Problem Definition

Medical experts regard a skin lesion that is nearly circular or elliptical in overall shape as more likely to be normal than not [4]. In view of this we have taken the ellipse to be a shape model for a normal skin lesion and to represent the most *regular* instance of the lesion shape. The more irregular the lesion border, the less likely it is to conform to the model. Figure (2) shows examples of lesion outlines, (a) regular and (b) irregular.

A "distance" between the model and a given lesion data is a measure of the lesion's irregularity. If irregularity is framed as a lack of predictability (see section 1), it is easy to see that the points on the circumference of an ellipse can be easily predicted whereas points belonging to an irregular shape are not easily predictable.

The lesion border is represented as a sequence of (1D) radial coordinates in a polar coordinate system centred at the centre of gravity of the lesion. The coordinates constitute an observation sequence $O = O_1, O_2, ..., O_M$ where O_i, i=1, ..., M is an i-th boundary point. A model λ describes a set of ellipses representing a normal skin lesion. The transition probability matrix a_{ij} constructed during training encodes a set of probabilities that the state S_i is followed by the state S_j for observations O_t (ellipse coordinates). If an unseen set of observations, O is presented, the probability P(O/ λ) will be greater if O corresponds to the model (i.e. it is an ellipse) than if it corresponds to an arbitrary shape. P(O/ λ) can be interpreted as the degree of conformity when O is generated using the knowledge λ. Our hypothesis is that $log(P(O|\lambda))$ should decrease with increase in border *irregularity*.

3.2 Choice of Probability Distribution for the Observation Sequence

The choice of a probability distribution was investigated using the quantile-quantile (Q-Q) plot method [11]. For the most appropriate distribution the data will have the largest linear correlation coefficient. The mean and standard deviation of the coefficients for the entire lesion data set were 0.977 (0.023) for Weibull distribution, 0.975 (0.025) for Gaussian distribution and 0.876 (0.066) for Gamma distribution. Figures (1 a, b and c) show the Q-Q plots for a typical benign skin lesion using Gamma distribution, Gaussian distribution and Weibull distribution respectively. It can be seen that both Gamma distribution Gaussian distribution have more outliers than the Weibull distribution.

4 Experiments

Experimental data consisted of 98 skin lesions of which of 16 were histologically confirmed cases of melanoma and the remaining 82 were benign lesions [12]. The radial coordinates corresponding to lesion boundary were extracted using a boundary modelling technique [13].

As the elliptical shape is likely to correspond to normal lesions (see 3.1), a Weibull distribution based HMM of the regular (elliptical) boundary was trained on a small set of ellipses, each represented by a sequence $y = y_1, y_2, ..., y_M$ of normalised radial coordinates. A model λ developed through this training is assumed to capture the essential characteristics of all similar y-s. Training employed an expectation maximisation method based on Baum-Welch algorithm [10]. The testing procedure was carried out by computing the log(P(O/ λ)) for the entire experimental data.

The experiments were repeated with the HMM trained using Gaussian probability distribution in place of Weibull distribution. The discriminatory power of the log-likelihood measure was assessed by performing the ROC analysis for both HMMs (i.e Weibull and Gaussian distribution trained). Lesions were arranged in order of decreasing log-likelihood measure, so that according to the hypothesis the most regular lesion would be first, the least regular would be last. The operating point (OP) was set for every value of the log-likelihood and sensitivity and specificity at this point was computed by counting how many benign (presumed regular) lesions had log-likelihood measure below the OP threshold and how many above.

One interesting question, not answered through the above experiments, was whether any of the computed measures corresponds to the human perception of the border irregularity of skin lesions. To this end an experimental survey

was carried out. 20 skin lesion outlines randomly selected from the full data set of which 20% represented malignant lesions, were given to 23 none medical experts. They were asked to rank the outlines based on their degree of irregularity. The level of agreement between their assessments was evaluated using rank correlation r_s based on Kendall coefficient of concordance W [14]. To test a "default" hypothesis, that irregularity simply depends on the magnitude of variations along the boundary, the standard deviation was computed for all the lesion outlines.

The Spearman coefficient of correlation was determined for each pair of the three irregularity estimations: the log-likelihood measure, the visual assessment (using the average ranking from the 23 subjects), and the standard deviation for the selected 20 lesions. Finally, the relationship between all three estimations was examined using multiple linear regression analysis.

5 Results and Discussion

The ROC analysis of the log-likelihood measure as a melanoma classifier showed 75% sensitivity and 76% specificity when the observation sequence was modelled using Gaussian probability distribution and 82% sensitivity with 82% specificity when using Weibull probability distribution. Standard deviation gave sensitivity of 77% and specificity of 78 %. Figure (3) shows the ROC plots for all three cases. On the basis of these results we have concluded that boundary sequences corresponding to abnormal lesions are less predictable (i.e. more irregular) when assessed using the Weibull probability distribution model than when using the Gaussian model of a regular (assumed normal) boundary. The departure from the Gaussian model of the regular lesion as measured by log-likelihood is no more discriminatory than a simple standard deviation of the boundary data.

In the experiments examining the perception of irregularity, the coefficient of concordance W was 0.886, indicating good agreement between all 23 subjects. The assessment of irregularity by the subjects correlated moderately well with both the Weibull based irregularity measure (Spearman correlation coefficient of 0.568) and standard deviation (0.466). This could be partially due to the small size of the dataset used in these experiments. Correlation was very poor between the Weibull based irregularity measure and standard deviation (0.229). The multiple linear regression analysis for all three tests showed equally poor correspondence (0.228). These results suggest that humans have similar way of assessing shape irregularity, but the human notion of irregularity may not be the same as measured by the indicators investigated in this paper. One possible interpretation of these results is that the features of a lesion boundary which makes it *look* irregular are not necessarily those associated with the irregularities present in abnormal lesions. A suitable computational irregularity measure may thus be a better abnormality predictor. This interpretation is only tentative because, first, the assessment of irregularity by clinicians is most likely to be different than those with no medical training; and second, they are drawn based on a small number of subjects (23). However, they are suggestive enough to encourage further work on this problem.

We have compared the result of the proposed measure with one of the best published melanoma predictors based on irregularity, the Indentation Irregularity Index (III) [12,15]. The comparison used the area under the ROC curve, which is a global measure commonly used to assess the overall predictive power of classification schemes. The III computed for a superset of the set of lesions used in our experiments has the area under ROC curve of 0.73 [12], whereas for the HMM based measure the area is 0.81. This indicates that the HMM-derived measure has a greater discriminatory power than the III index. Using Wicoxon statistic [16], both the III index and our measure were significantly better than chance with $P < 0.0001$. The comparison of the methods using the same statistic showed that our method was slightly more predictive than III index, with Z=1.0373, $P < 0.15$.

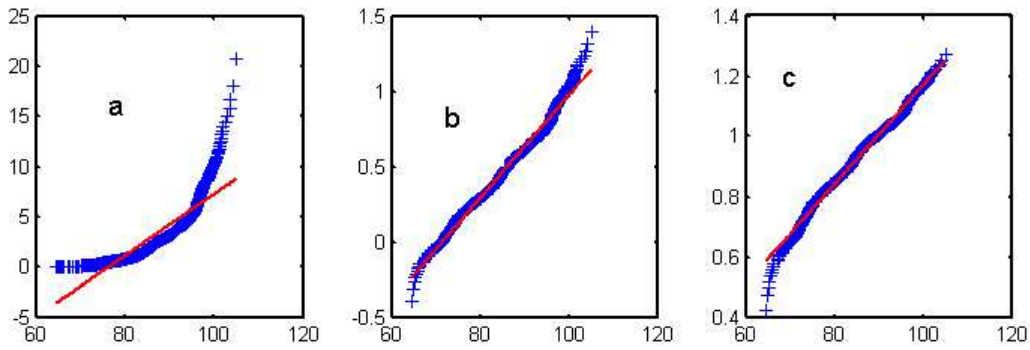

Figure 1. Q-Q plot: (a) Gamma distribution, (b) Gaussian distribution, (c) Weibull distribution

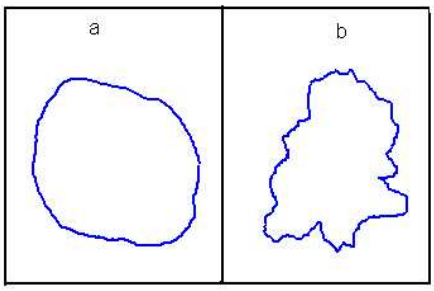

Figure 2. Sample Lesion Outlines:(a) regular, (b) irregular

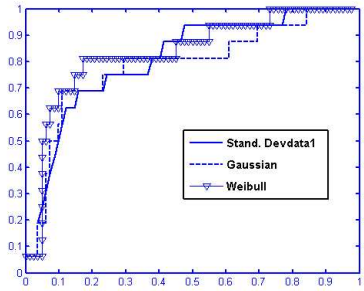

Figure 3. ROC curves for the three measures - as melanoma predictor

The irregularity measure discussed in this paper uses as a criterion the lack of predictability. Our earlier work proposed four other criteria of irregularity, namely lack of compressibility, asymmetry, lack of rule, and deviation from rule [9]. It will be interesting to investigate a combined irregularity measure which incorporates these additional criteria.

6 Conclusion

In this paper we have proposed a new measure of border irregularity for pigmented skin lesions based on Hidden Markov Models. The measure has been devised to quantify one of the attributes of irregularity, namely a lack of predictability. We have demonstrated that regular lesion boundaries can be modelled using a Hidden Markov Model assuming Weibull distribution. Irregular boundaries do not conform to this model, which results in the decrease of the log-likelihood of the boundary being represented by the model. ROC analysis of the log-likelihood as a malignancy predictor gave 82% of both sensitivity and specificity. This result shows that the model captures some distinguishing features in the boundary of malignant lesions and thus can contribute to lesion classification.

References

1. M. Keefe, D. Dick & R. Wakeel. "A study of the value of the seven point checklist in distinguishing benign pigmented lesions from melanoma." *Clin. and Exp. Derm.* **15**, pp. 167 – 171, 1990.
2. J. D. Morris-Smith. "Characterisation of the appearance of pigmented skin lesions." *Ph.D. thesis, The University of Birmingham, U.K.* 1996.
3. R. Mackie. "Malignant melanoma, a guide to early diagnosis." *Tech. Rep., University Glasgow* 1989.
4. E. Claridge, P. Hall, M. Keefe et al. "Shape analysis for classification of malignant melanoma." *J. Biomed. Eng.* **14(3)**, pp. 229–234, 1992.
5. W. Stoecker, R. H. Moss, F. Ercal et al. "Nondermatoscopic digital imaging of pigmented lesions." *Skin Research and Technology* **1**, pp. 7–16, 1995.
6. P. Hall, E. Claridge & J. Smith. "Computer screening for early detection of melanoma - is there a future." *British Journal of Dermatology* **132**, pp. 325 – 338, 1995.
7. D. Gutkowicz-Krushin, M. Elbaum, P. Szwaykowski et al. "Can early malignant melanoma be differentiated from atypical melanocytic nevus by in vivo techniques ?" *Skin Research and Technology* **3**, pp. 15–22, 1997.
8. T. K. Lee, D. I. McLean & M. S. Atkins. "Irregularity index: A new border irregularity measure for cutaneous melanocytic lesions." *Medical Image Analysis* **7(1)**, pp. 47–64, 2003.
9. B. Aribisala. "Computing irregularity for features in medical images." *Thesis Proposal, The University of Birmingham* May 2003.
10. L. R. Rabiner. "A tutorial on Hidden Markov Models and selected applications in speech recognition." *Proceedings of the IEEE* **77(2)**, pp. 275 – 286, 1989.
11. J. M. Chambers, W. Cleveland, B. Kleiner et al. *"Graphical Methods for Data Analysis." Wadsworth* 1983.
12. T. K. Lee & E. Claridge. "Predictive power of irregular border shapes for malignant melanomas." *Skin Research and Technology* **11(1)**, pp. 1–8, 2005.
13. E. Claridge & A. Orun. "Modelling of edge profiles in pigmented skin lesions." *Medical Image Understanding and Analysis 2002* pp. 53–5, 2002.
14. W. Hays. *"Statistics*, chap. 19." *Holt, Rinehart and Winston* pp. 836–847, 1988.
15. T. K. Lee & M. Atkins. "A new shape measure for melanocytic lesion." *Proceeding of Medical Image Understanding and Analysis, London,U.K.* pp. 25 – 28, 2000.
16. J. Fogarty, R. Baker & S. Hudson. "Case studies in the use of roc curve analysis for sensor-based estimates in human computer interaction." *To Appear, Proceedings of the ACM Conference on Human Factors in computing Systems* 2005.