Chapter 5: Results

5.1 Introduction

In this chapter the results of the research project are presented. Initially the reliability of the SIAscopic features that were defined and refined for the project was assessed using Kappa statistics. Following on, the useful features that were identified were used to develop predictive models for the diagnosis of malignant melanoma using logistic regression and classification tree analysis techniques. Once the models were constructed, classification tables were produced and specificity and sensitivity analyses were performed that included receiver-operating characteristic (ROC) curves.

In addition to the 'traditional' statistical predictive models that were developed from the datasets, a scoring system was designed. Scoring systems or checklists in the appraisal of pigmented skin lesions are popular with clinicians. They allow rapid and focused assessment of the lesion with clear guidance on whether to observe/reassure the patient or to refer them for either an expert opinion or excision biopsy. The Revised 7-point Checklist is a scoring system devised for General Practitioners in order to flag up lesions for the attention of the Dermatologist [MacKie, 1990]. Any lesion that scores greater than 2 points is deemed to be suspicious and warrants further investigation. In addition to this clinical checklist, there are three dermatoscopic checklists that are in common use that were recently compared and evaluated in a large-scale clinical study [Soyer *et al.,* 2001]. The first of these is the 'ABCD' method of skin surface microscopy devised by Stolz *et al.* (1994) where the acronym stands for Asymmetry, Borders, Colours & Dermatoscopic structures. Each of these criteria are assessed and multiplied by a given weight factor that generates a total dermatoscopy score (TDS). TDS values greater than 5.45 are strongly suggestive of melanoma, values between 4.8 and 5.45 are suspicious of melanoma and a value less than 4.75 indicates a benign melanocytic lesion. The second of the dermatoscopic checklists is Menzies' method [Menzies *et al.,* 1996b]. This checklist requires two negative features to be absent *and* one positive feature to be present to diagnose a melanoma. The final method is the dermatoscopic 7-Point Checklist [Argenziano *et al.,* 1998]. This method has similarities to MacKie's Revised 7-Point checklist (1990) in that there are three major criteria and four minor criteria that score two points and one point respectively. A total score of three or greater indicates that the lesion is likely to be a melanoma. The criteria were sorted into major and minor categories on the basis of the odds ratio of that predictor – any predictor with an odds ratio greater than five was classed as a major criterion and those with an odds ratio less than five were classed as minor. Comparison of all three methods reveals strikingly similar sensitivity and specificity results – approximately 85% and 70% respectively [Soyer *et al.,* 2001; Argenziano, 2001]. These methods have the advantage of allowing rapid and objective assessment of the lesion rather than just 'eyeballing' it and arriving at a diagnosis. The disadvantage of these methods are

that they are constraining and do not take account of the global appearance of the lesion. As a result the specificity of these methods is lower than those pattern analysis techniques [Soyer *et al.*, 2001; Argenziano, 2001].

5.2 Intra- & Inter-Observer Agreement of SIAscopy Features

Tables 5.1 & 2 and figures 5.1 & 2 show the results of the intra- and inter-observer agreements of the SIAscopy features respectively. The strength of agreement is measured using the kappa statistic that was described in section 3.5

Table 5.1 Kappa Scores for ***Intra***-Observer Agreement

Feature	Kappa Score	Standard Error	95% Confidence Intervals	
			Lower	Upper
Blood Displacement	0.827	0.050	0.7290	0.9250
Dermal Melanin	0.818	0.050	0.7200	0.9160
Blood Displacement with Blush	0.809	0.061	0.6894	0.9289
Dermal Melanin Globules	0.798	0.053	0.6941	0.9019
Asymmetry	0.636	0.067	0.5047	0.7673
Collagen Holes	0.631	0.069	0.4958	0.7662
Melanin Globules	0.602	0.100	0.4060	0.7980
Biaxial Symmetry	0.586	0.077	0.4351	0.7370
Erythematous Blush	0.501	0.076	0.3520	0.6500
Symmetry in One Axis	0.399	0.091	0.2206	0.5774
Blood Globules	0.304	0.069	0.1688	0.4392

Table 5.2 Kappa Scores for ***Inter***-Observer Agreement

Feature	Kappa Score	Standard Error	95% Confidence Intervals	
			Lower	Upper
Dermal Melanin	0.879	0.026	0.8280	0.9300
Blood Displacement	0.783	0.035	0.7144	0.8516
Dermal Melanin Globules	0.671	0.040	0.5926	0.7494
Blood Displacement with Blush	0.664	0.046	0.5738	0.7542
Asymmetry	0.608	0.043	0.5237	0.6923
Biaxial Symmetry	0.589	0.046	0.4988	0.6792
Collagen Holes	0.522	0.048	0.4279	0.6161
Erythematous Blush	0.496	0.045	0.4078	0.5842
Melanin Globules	0.362	0.072	0.2209	0.5031
Symmetry in One Axis	0.328	0.055	0.2202	0.4358
Blood Globules	0.261	0.062	0.1395	0.3825

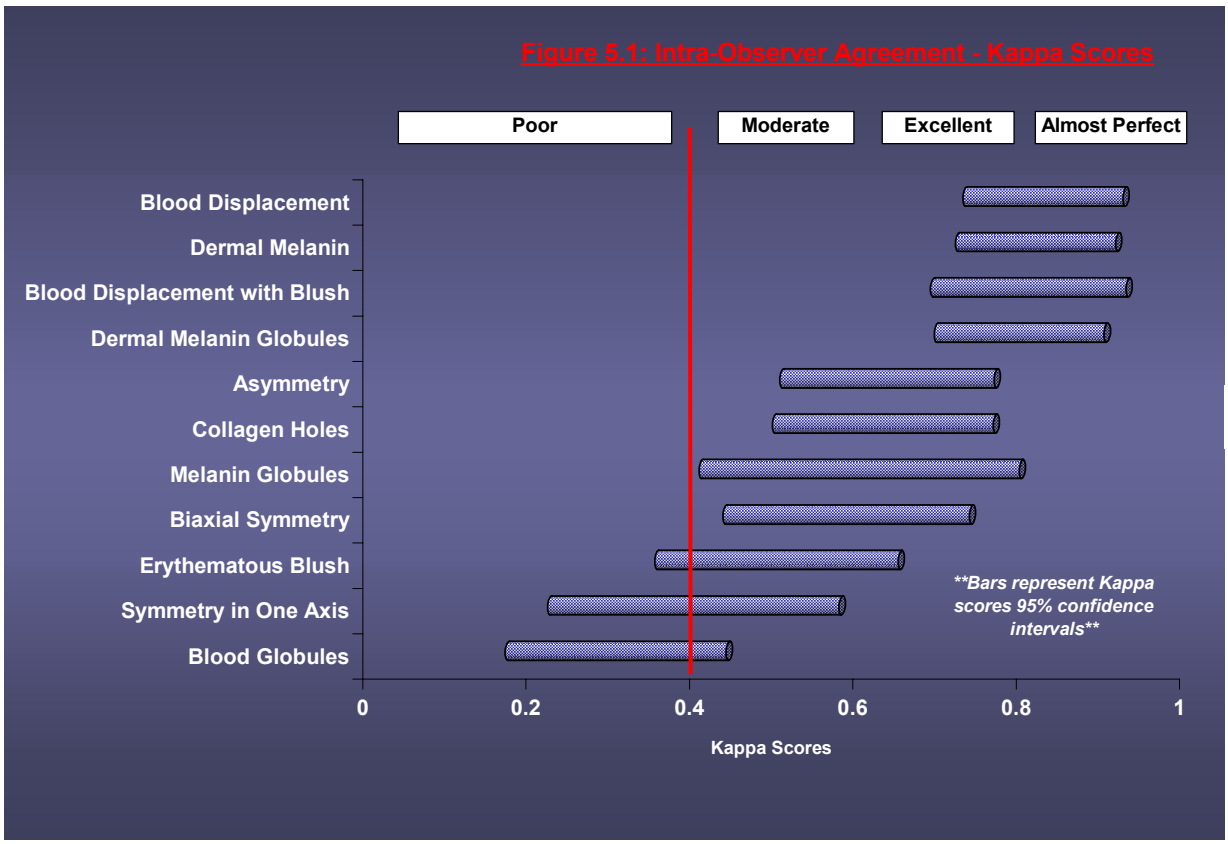


Figure 5.1: Intra-Observer Agreement - Kappa Scores
Poor
Moderate
Excellent
Almost Perfect
Blood Displacement
Dermal Melanin
Blood Displacement with Blush
Dermal Melanin Globules
Asymmetry
Collagen Holes
Melanin Globules
Biaxial Symmetry
Erythematous Blush
Symmetry in One Axis
Blood Globules
Bars represent Kappa scores 95% confidence intervals
0
0.2
0.4
0.6
0.8
1
Kappa Scores

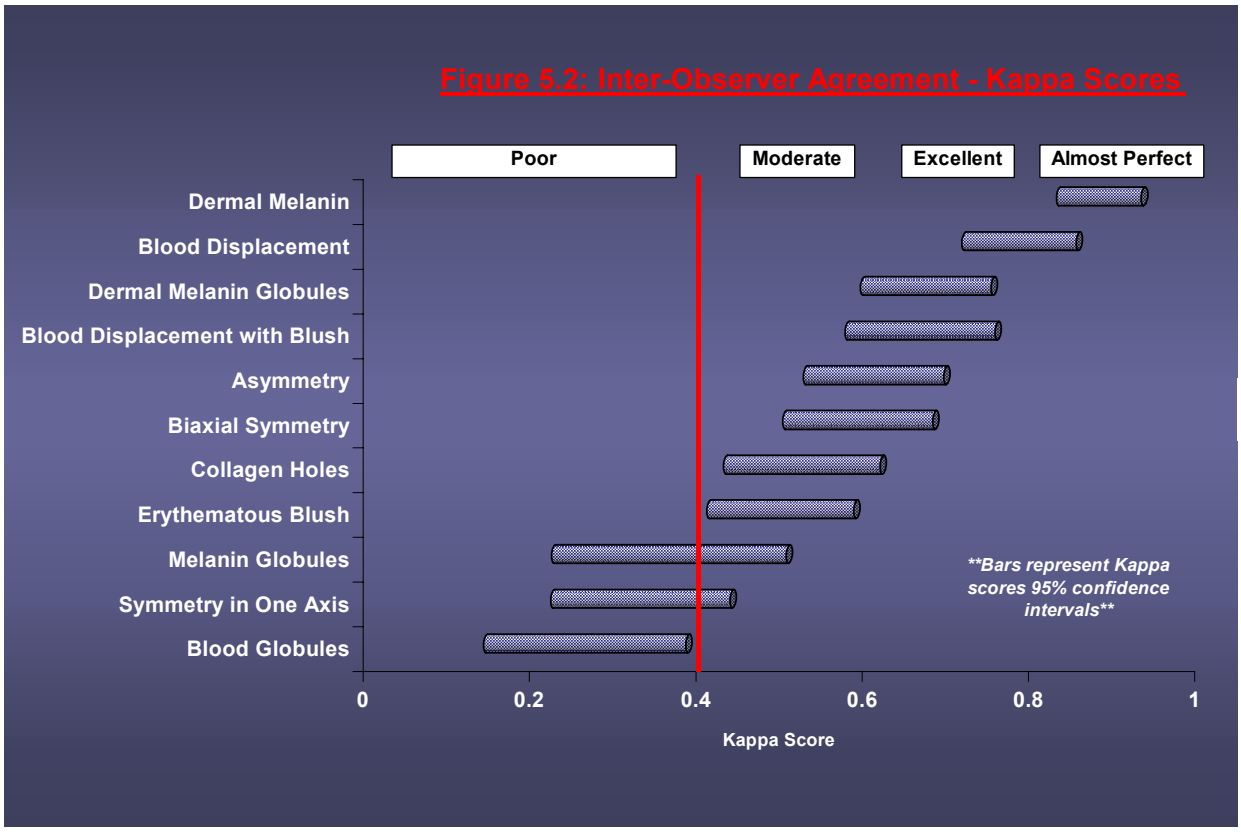


Figure 5.2: Inter-Observer Agreement - Kappa Scores
Poor
Moderate
Excellent
Almost Perfect
Dermal Melanin
Blood Displacement
Dermal Melanin Globules
Blood Displacement with Blush
Asymmetry
Biaxial Symmetry
Collagen Holes
Erythematous Blush
Melanin Globules
Symmetry in One Axis
Blood Globules
Bars represent Kappa scores 95% confidence intervals
0
0.2
0.4
0.6
0.8
1
Kappa Score

5.2.1 Discussion

The tables for intra- and inter-observer agreement demonstrated consistent results for the SIAscopy features. Dermal melanin and blood displacement achieved the highest kappa scores that are interpreted as ‘almost perfect’ [Kianifard, 1994]; this is not surprising as they are the simplest features to identify. The mental process of detecting dermal melanin runs along the lines of ‘is there any blue on the screen where the lesion is’. Similarly, the process of detecting blood displacement runs along the lines of ‘is there white on the screen where the lesion is’. There are some caveats as was discussed in section 3.2.1.2 but, in general, there is little margin for subjectivity in the decision process. These two features are thus considered to be very repeatable and reproducible and can be identified with high objectivity.

The next three features that registered high kappa scores were blood displacement with blush, dermal melanin globules and asymmetry. These values are interpreted as demonstrating ‘excellent’ or ‘almost perfect’ agreement [Kianifard, 1994]. Their kappa scores were lower than dermal melanin and blood displacement as the decision process to confirm or refute their presence involves a greater degree of complexity. For all of these features there are at least two variables to be considered simultaneously before arriving at a judgment. As a result there is a greater level of subjectivity that can influence the outcome of that decision.

It is noted that the top five features for intra- and inter-observer agreement are the same and that the scores in the two tables are in the same category according to Kianifard (1994). This would imply that the definitions for these five features are the most robust. If this had not been the case it could be speculated that the definitions for the features were too difficult to interpret, that there was some level of ambiguity to the definitions that allowed for different interpretation or that the definitions were sufficiently lax or complicated that the level of subjectivity introduced was too great to render them useless.

These features had been highlighted before the analysis began as being likely to be very useful predictors for malignant melanoma. In addition, these features are the ones that present the newest information to the clinician in terms of underlying pathophysiology of the lesion. The presence of dermal melanin and displacement with blush are the two most novel features. Using a dermatoscope, it is possible to obtain indirect evidence that there is dermal melanin from the presence of blue-grey veil (section 2.3.1.1). However, the examiner may perceive the colour blue via several optical pathways that include thinning of the papillary dermis, hyperkeratosis, deoxyhaemoglobin and thrombus [Cotton, 1998]. Moreover, the presence of blue may be masked by excess epidermal melanin [Cotton, 1998]. Thus, this is not a reliable or repeatable sign due the optics of the lesion alone. Dermal melanin can be reliably and reproducibly identified in the SIAscopy images. It is expected that this will prove to be an effective predictor of malignant melanoma, as the presence of dermal melanin will help to discriminate between this and a compound naevus where the dermal melanocytes have

atrophied or undergone maturation. Using a dermatoscope, it also possible to appreciate an increased vascularity within the lesion [Menzies *et al.*, 1996a]. However, the vascularity of the lesion is usually masked by the overlying epidermal melanin. In addition, displacement of blood supply as indicated by scarring is a gross feature in skin surface microscopy, usually only seen in advanced or regressed melanomas [Menzies *et al.*, 1996a, Soyer *et al.,* 2001]. The presence of blood displacement with erythematous blush can be reliably and reproducibly detected from the SIAscope images. It is believed that the combination of these features represents the underlying process of invading melanoma displacing the papillary blood supply and exciting a local immune response that produces peripheral vasodilatation of the microvasculature (section 3.2.1.1). Finally, asymmetry appears to be a reliable and reproducible SIAscopy feature. In assessing lesions both clinically and dermatoscopically, it has been shown that asymmetry is a strong predictor of melanoma [Fitzpatrick et al., 1988; MacKie, 1990; Stolz *et al.*, 1994; Menzies *et al.,* 1996b] and it is anticipated that asymmetry as determined by the SIAscope will also be a strong predictor of melanoma. In addition, the total melanin SIAgraph provides an intuitive method of segmenting the lesion into the melanin-containing region. The presence of asymmetry is produced by the haphazard and uncontrolled spread of melanin-containing melanoma cells and it is expected that asymmetry determined by SIAscopy will reduce the number of false positives that are caused by non-melanocytic irregular lesions.

The next three features that were repeatable and reproducible were collagen holes, biaxial symmetry and erythematous blush. As a group they scored in the moderate to excellent category indicating a fair amount of subjectivity in their identification.

Despite having only one layer of complexity to consider in the identification of collagen holes,

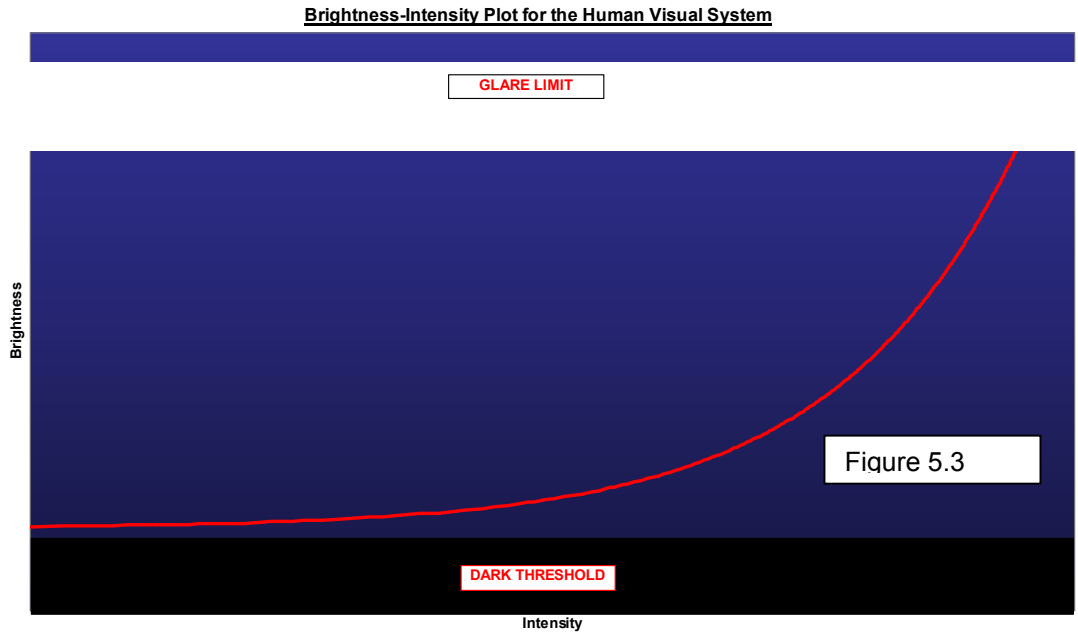


Figure 5.3

it still remained a difficult task. A small part of the reason for this may lie in the limited axial resolution of the SIAscope (section 3.7.3.3). However, the main subjectivity error in identifying the collagen holes may arise from the presentation of the collagen SIAgraphs and the intensity-brightness response curves in the human visual system - Figure 5.3. The curve describes a logarithmic function where the gradient is steep for higher intensities than for lower ones [Umbaugh, 1998]. In contrast to the human visual system, the cathode ray tube monitor outputs light intensity in a linear manner. This information has two implications. First, whilst a monitor can output 255 different shades of grey, the human visual system can only detect approximately 100 [Umbaugh, 1998]. Second, at lower levels of light, the light intensity has to approximately double before a change in brightness is perceived and at higher levels of intensity, the human visual system rapidly approaches the 'glare limit' with small changes in light intensity. The collagen images mainly contain data whose values are at either extremes of the output scale for the monitor. Thus the pixel values are perceived as being either very black or very white and as a result subtle collagen holes are difficult to identify.

In contrast to asymmetry, biaxial symmetry had a reasonable level of subjectivity in its identification according to the kappa scores. This may be explained by the considering the logic involved in ascertaining biaxial symmetry and asymmetry when considering the three variables of border, content of the lesion and axis of the lesion. Asymmetry is a logical 'OR' operation as follows, '**NOT** (Symmetry in border) **OR NOT** (asymmetry in content) = Asymmetry.' Biaxial symmetry, on the other hand, is a logical 'AND' operation, 'Symmetry in border **AND** Symmetry in Content **AND** Two Axes at 90^{o} to each other = Biaxial Symmetry.' Symmetry in border and content of the lesion is described in terms of biological variability, as no single lesion will be perfectly symmetrical for either of these parameters. This meant that a certain amount of thresholding had to take place to implement the decision process and, as a result, subjectivity crept in. As three variables had to be considered for the diagnosis of biaxial symmetry this served to compound the subjectivity and thus lessen the kappa scores for intra- and inter-observer agreement. Interestingly, symmetry in one axis had a very poor intra- and inter-observer agreement. Whilst it is difficult to prove, one assumption may be that this observation arose out of the tendency for human observers in general to favour extremes of cases i.e. total symmetry or total asymmetry.

In comparison to blood displacement, erythematous blush showed only a 'moderate' agreement in intra- and inter-observer agreement. Indeed, the 95% confidence intervals for intra-observer agreement strayed into the 'poor agreement' category according to Kianifard (1994). The reason for this level of subjectivity in determining this feature is likely to be due to the change in micro architecture from normal skin to that containing a melanocytic lesion. Mooi & Krausz (1992a) state that the rete ridges are elongated in benign naevi. Concurrently, the papillary dermis is also elongated at the ridges, as is the vasculature that passes vertically upwards to dermo-epidermal junction. As a result, any light travelling vertically through the

papillary dermis of a melanocytic lesion will functionally encounter more haemoglobin than that that passes through normal skin. Thus, melanocytic lesions tend to display a diffuse increase in vascularity compared to normal skin when viewed with a SIAscope. This is considered to be artefact as it is very unlikely that it is brought about by vasodilatation or neoangiogenesis that is seen in melanomas. The observer has to distinguish the presence of erythematous blush from this artefact by a process of thresholding. It is likely that a human observer will be relatively inconsistent in this process leading to the relatively low kappa scores. However, it has already been shown that the combination of erythematous blush with blood displacement has excellent intra- and inter-observer agreement. It is assumed that this occurs for two reasons: first, the identification of blood displacement is very simple and second, erythematous blush tends to occur in combination with blood displacement anyway (though the converse is not true).

Melanin globules, blood globules and symmetry in one axis consistently scored poorly. The reasons why this might be for the latter feature have been discussed above. Of all the lesions analysed 293 (84.2%) contained blood globules and 323 (92.8%) contained melanin globules. Thus, it can be stated that these features were more or less ubiquitous in the dataset examined. Therefore, it is reasonable to assume that any attempt to discriminate the presence or absence of globules would result in large errors between observers as subtle variations in thresholding will result in large sections of data being included or excluded in each category. Given that these features are so ubiquitous it is also reasonable to assume that their inclusion in an attempt to discriminate between melanoma and non-melanoma will not be very useful.

In conclusion to this section, it was shown that the majority of SIAscopy features had good intra- and inter-observer agreement and that they could be used in developing a predictive model for diagnosing melanoma. As a result of the kappa statistic analysis it was also decided that blood, melanin, dermal melanin globules and symmetry in one axis would not be used in the development of logistic regression and classification tree analysis models.

5.3 Developing Predictive Models Using Logistic Regression

5.3.1 The Model-Building Dataset

The model-building dataset was collected over a 12-month period from January 2000 to January 2001. Table 5.3 shows the diagnostic case-mix of all the lesions that were collected. The original dataset comprised of 362 lesions and sets of images. However, 14 (4.0%) sets of images were corrupted and considered to be uninterpretable; this was mostly due to light-leakage artefact. Thus the remaining dataset comprised of 348 lesions (311 patients – 200 female & 111 male), including 52 melanomas (1 in 6.70 lesions). Of the melanomas, 6 were *in situ*, 11 had a Breslow thickness less than 0.76mm and in total 28 lesions had a Breslow thickness of 1.0mm or less (range *in situ* to 6.0mm). The median & mean Breslow thicknesses

were 1.0mm & 1.52mm respectively, and the subtypes included 41 superficial spreading melanomas, 9 nodular and 2 acral lentiginous melanoma. Given this case-mix and the large proportion of 'thin' melanomas it was decided that this dataset was representative and would be functional for devising a predictive model using logistic regression analysis.

For the purposes of building the model, lesions were classified as melanoma or non-melanoma. This is highly artificial and makes redundant other clinically useful information such as melanoma subtype. In addition, the non-melanoma subgroup includes pigmented basal cell carcinomas. These lesions require excision as they are malignant and, although they do not have the metastatic potential of melanomas, can be locally destructive and aggressive. This is an important point to be considered when qualifying the predictive model and will be discussed further in the conclusions section of this thesis.

Table 5.3 Diagnostic Case-mix for Model-Building Dataset

Diagnosis	**Count**
Melanoma	52
- Superficial Spreading	*41*
- Nodular	*9*
- Acral Lentiginous	*2*
Common Naevi (Compound, Junctional & Intradermal)	185
Dysplastic Naevi	7
Blue Naevi	12
Spitz Naevi	7
Seborrheic Keratoses	29
Lentigo	9
BCC	21
Dermatofibroma	8
Mixed Naevi	2
Haemangioma	2
Others	14
Total	**348**

5.3.2 Univariate Analysis

Initially, each of the SIAscopy features were analysed for their predictive power on an individual basis (Table 5.4). Analysis showed that no single feature was both highly specific and highly sensitive for melanoma. The presence of dermal melanin was highly sensitive for

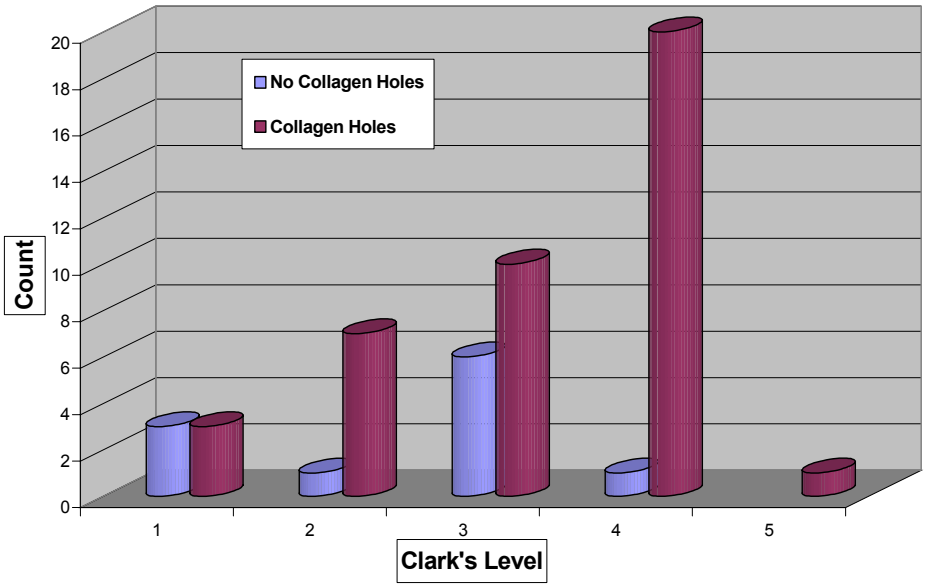
Table 5.4 Results of Univariate Analysis

Predictor	True Positives (max = 52)	Sensitivity	95% Confidence Intervals		True Negatives (max = 296)	Specificity	95% Confidence Interval	
			Lower	Upper			Lower	Upper
Dermal Melanin	50	0.962	0.870	0.989	168	0.568	0.511	0.623
Blood Displacement With Erythematous Blush	33	0.635	0.499	0.752	251	0.848	0.803	0.884
Collagen Holes	41	0.788	0.660	0.878	219	0.740	0.687	0.787
Asymmetry	40	0.769	0.639	0.863	184	0.622	0.565	0.675
Blood Displacement	39	0.750	0.618	0.848	208	0.703	0.648	0.752
Erythematous Blush	39	0.750	0.618	0.848	194	0.655	0.600	0.707

melanoma in that fifty of the fifty-two lesions demonstrated it. This means that, in this dataset, an absence of dermal melanin means that the lesion is very unlikely to be a melanoma. However, the specificity of dermal melanin is low, with a value of 0.568. This means that the presence of dermal melanin was found in almost half of the benign lesions. Therefore, when considering dermal melanin alone, an absence would be considered diagnostic of a benign lesion but its presence would not clinch the diagnosis of a melanoma. An important observation here is the number of melanomas that displayed this feature. This is, in part, surprising as six of them were *in situ* lesions that are, by definition, confined to the epidermis. However all of these demonstrated dermal melanin. This may be explained by the fact that, as a result of processing the lesion for microscopy, not all the melanoma is examined by the histopathologist. It is possible that the SIAscope had detected very small areas of invasion that were not available to the Histopathologist. Furthermore, *in situ* melanomas can demonstrate pigmentary incontinence and melanophages (chapter 1) and it is likely that this is the source of the dermal melanin that is detected by the SIAscope [Mooi & Krausz, 1992a].

In contrast, the presence of blood displacement with erythematous blush was highly specific for melanoma but had a poor sensitivity. The presence of this feature meant that the lesion

Figure 5.4: Melanomas With & Without Collagen Holes by Clark's Level



was highly unlikely to be benign, yet the absence of this feature was not able to rule out the diagnosis of melanoma. However when analysed alone, neither erythematous blush nor blood displacement had a good specificity or sensitivity. This does appear to make sense clinically as several benign lesions, especially blue naevi, display blood displacement yet no erythematous blush. In contrast, many benign naevi display erythematous blush but no blood displacement. It is likely that it is the presence of the two features together that is indicative of invasive melanoma and the underlying pathophysiological processes occurring within it. Thus, the presence of either one alone is a non-specific finding.

The SIAscopy features of asymmetry and collagen holes had neither a good sensitivity nor a good specificity. This is likely to be due in large part to the reduced reliability and reproducibility in the identification of these features when compared to dermal melanin and blood displacement with erythematous blush. In addition, it is suggested that collagen holes are a marker of advanced primary melanoma. This is likely to be because in order to produce a collagen hole the melanoma has to have invaded to Clark's level III or greater. Evidence for this can be obtained from the dataset. If all 52 of the melanomas are analysed by Clark's level it is apparent that the majority of the lesions (83.80%)that are Clark's level III or greater display collagen holes (Figure 5.4). In addition, statistical analysis for trend using Kendall's Tau-b statistic [Bland, 1997] shows a significant difference between the two groups (Tau-b = 0.288, SE = 0.108, p=0.014). Therefore, collagen holes are indicative of a locally advanced melanoma yet the dataset has a median Breslow thickness of 1.0mm that means that the majority of the lesions were displaying early local disease.

5.3.3 Constructing the Logistic Regression Models

Logistic Regression analysis was performed on this set of lesions using SPSS for Windows v9.0 (SPSS Inc., USA). Two experimenters performed feature analysis on the data so that the Kappa scores for inter-observer agreement could be determined as shown in section 5.2. However, the dataset that was used for the statistical analysis in the remainder of this chapter was that produced by the author. The reason for this was that the second experimenter, whilst very knowledgeable on the subject of melanoma, is not a practicing clinician.

Table 5.5 Goodness of Fit Statistics for 'SIAscopy' Logistic Regression Model

Step	Variable	Step Change Deviance	Significance	Model Deviance
0	Constant	-	-	293.511
1	Blood Displacement with Erythematous Blush	49.752	< 0.001	243.759
2	Dermal Melanin	27.382	< 0.001	216.378
3	Asymmetry	14.517	< 0.001	201.861
4	Collagen Holes	4.912	0.027	196.984

Forward selection and backwards deletion stepwise regression methods were initially performed. Stepwise methods were initially chosen as this was an exploratory project where no previous research existed on which to base theories for testing (section 3.4.2)[Field, 2000]. Both models constructed were identical. Table 5.5 summarises the goodness-of-fit statistics of the forward selection model. From the tables it can be seen that four features were included in the model, namely blood displacement with erythematous blush, dermal melanin, asymmetry and collagen holes. The inclusion of each variable produced a significant change in deviance and the model chi-square statistic is highly significant. However, the model deviance is also significant producing the situation where the model variables produce a significant improvement in prediction of the outcome variable but the model is still a relatively poor fit of the data (section 3.4.3.2). Table 5.6 Summarises the statistics for each feature in the equation.

Table 5.6 Summary Statistics for SIAscopy Logistic Regression Model

Feature	Regression Coefficient, β	Wald Statistic	Significance	Pseudo R-Value	Exp(β) (Odds Ratio)	95% CI for Exp(β)	
						Upper	Lower
Constant	-5.1773	46.8937	< 0.001	-	-	-	-
Blood Displacement with Erythematous Blush	1.1486	9.4429	0.0021	0.1592	3.1539	1.5159	6.6518
Dermal Melanin	2.3240	8.7885	0.0030	0.1521	10.2166	2.1980	47.4892
Asymmetry	1.1955	9.2517	0.0024	0.1572	3.3052	1.5298	7.141
Collagen Holes	0.9145	4.6973	0.0320	0.0959	2.4954	1.0914	5.7054

All of the regression coefficients and pseudo R-values are positive for the SIAscopy features. This means that the presence of these features increases the odds of the lesion being a melanoma. If any of the features decreased the odds of the lesion being a melanoma then there would be serious concerns raised about the model, the dataset or the feature analysis. Each of the features has a significant Wald statistic so each one is significantly different from zero and is making a contribution to the model in predicting melanoma. The largest change in odds is with the presence of dermal melanin. This can be explained by the histopathology where early invasive melanoma continues to produce melanin whereas the intradermal melanocytes of a compound or intradermal naevus can cease to do so [Mooi & Krausz, 1992a]. Asymmetry and blood displacement with erythematous blush equally increase the odds of the lesion being a melanoma. It may be expected that asymmetry would have a greater effect than the model suggests, given that such weight is given to this finding clinically and dermatoscopically [Fitzpatrick et al., 1988; Stolz *et al.*, 1994; Menzies *et al.,* 1996b]. However, many lesions displayed asymmetry (152 lesions – 43.7%) and it is likely that the lack of specificity of this feature decreases the size of the change in odds when it is present. Collagen holes have the smallest odds ratio value and this is may be related to the fact that this predictor is related to advanced primary melanoma rather than early local disease (section 5.3.2). In addition to this, the lower confidence interval for the odds ratio is close to 1 (1.091) indicating that there may be a very weak relationship between collagen holes and melanoma. It is also likely that better predictors of melanoma than collagen holes could be found when the dataset is further analysed using the clinical data.

The model includes neither blood displacement nor erythematous blush as single and separate predictors as they are non-specific findings (section 5.3.2). In addition to the blood SIAgraph features, biaxial symmetry was also excluded from the model. This was a surprising finding, as it was expected that this would have a significant negative correlation with

melanoma i.e. have a negative β value. The reason that it didn't is unclear but may be a result of the fact that only a small proportion of all the lesions displayed biaxial symmetry (90 lesions – 25.8%).

Table 5.7 Odds Table from Logistic Regression

Feature				Log odds	Odds for Melanoma	Odds ratio	Probability for Melanoma	Sensitivity	Specificity
DM	**AS**	**DB**	**CH**						
0	0	0	0	-5.177	0.006	0.032	0.006	1.000	0.000
0	0	0	1	-4.263	0.014	0.080	0.014	1.000	0.355
0	0	1	0	-4.029	0.018	0.101	0.017	1.000	0.372
0	1	0	0	-3.982	0.019	0.106	0.018	1.000	0.378
0	0	1	1	-3.114	0.044	0.253	0.043	0.981	0.541
0	1	0	1	-3.067	0.047	0.265	0.044	0.981	0.544
1	0	0	0	-2.853	0.058	0.328	0.055	0.962	0.564
0	1	1	0	-2.833	0.059	0.335	0.056	0.904	0.672
1	0	0	1	-1.939	0.144	0.819	0.126	0.904	0.676
0	1	1	1	-1.919	0.147	0.836	0.128	0.865	0.736
1	0	1	0	-1.705	0.182	1.035	0.154	0.827	0.770
1	1	0	0	-1.658	0.191	1.085	0.160	0.769	0.824
1	0	1	1	-0.790	0.454	2.583	0.312	0.673	0.861
1	1	0	1	-0.743	0.476	2.707	0.322	0.500	0.932
1	1	1	0	-0.509	0.601	3.420	0.375	0.462	0.949
1	1	1	1	0.405	1.500	8.536	0.600	0.000	1.000

Where:
0 = absence of feature
1 = presence of feature
DM = Dermal Melanin
AS = Asymmetry
DB = Blood Displacement with Blush
CH = Collagen Holes

Where:

Probability = Odds/(1+Odds)
Odds Ratio = Post-test / Pre-test Odds

The odds and probabilities generated from this model are shown in table 5.7. Specificities and sensitivities for these data are calculated and plotted as an ROC curve (figure 5.5). The area under this curve is 0.88 (95% confidence intervals: 0.838 – 0.921) and this is significantly different from 0.5 ($p < 0.001$). An area of 0.88 means that a randomly chosen lesion from the

melanoma group has a higher odds value than a randomly chosen lesion from the non-melanoma group 88% of the time [Campbell & Machin, 1999]. There are tied points along the length of the ROC curve so the maximal area is also calculated whose value is 0.908.

An analysis of the studentised residuals revealed that five cases lie outside 3 standard deviations and required further investigation. All of the lesions with large residual values were melanomas and the two with the largest studentised residual values were the only two that did not display dermal melanin on the SIAscope. The first lesion was a superficial spreading melanoma, Breslow thickness 0.4mm that measured only 6mm in diameter. In addition to having an absence of dermal melanin, this lesion did not display blood displacement with erythematous blush or collagen holes. These features were double-checked to ensure that the data had not been incorrectly entered. The second lesion also failed to display dermal melanin and was a small (4mm diameter) superficial spreading melanoma on the leg of a 67 year-old female. The histopathological diagnosis of this lesion was initially uncertain and slides of the specimen were sent for external review in this country and abroad. The consensus of expert opinion was unequivocally that of a melanoma and, as a result, it was felt that this lesion could not be removed from the dataset on the grounds of being initially difficult to diagnose histopathologically. Three further cases included two early superficial spreading melanomas and a thin nodular melanoma (Breslow 1.0mm). Whilst these three cases displayed dermal melanin, they were symmetrical in shape and content and did not display blood displacement with erythematous blush. This is surprising in the latter case, though further review of the images confirms a greatly increased vascularity that is often seen with these lesions [Mooi & Krausz, 1992a]. A review of all of these cases found that the data had been entered correctly and the histopathological diagnosis was not in doubt and therefore the cases must be included in the dataset. As was stated in section 1.3.2.2, there is not a single architectural feature that is always present that diagnoses melanoma histopathologically but instead a constellation of features that exclude benign naevi. This is borne out by the analysis of SIAscopy features. It is possible that invasive lesions not displaying dermal melanin contain malignant cells that, through the stepwise process of malignant transformation (section 1.3.2.3), have lost the ability to produce melanin. It was not possible to investigate this further on the dataset. Assessment of the influence of these outliers using Cook's distance and DFBeta scores reveals that they do not exert undue influence on the model.

This model was analysed for collinearity using Eigenvalue statistics and condition indices [Field, 2000]. All of the values obtained for the condition indices were similar and the uncentred cross-product matrix was well conditioned indicating that there was insignificant collinearity affecting the model.

The next step was to determine whether this model could be improved by the addition of the clinical data that was collected from the patient with the SIAscope images of the lesion and whether this model was applicable in subset analysis such as that with only 'thin' melanomas.

5.3.4 Adding Clinical Data

In adding clinical data, it was necessary to ensure that the information was reproducible and objective. For this reason, information about the lesion in the form of the Revised Seven-Point Checklist [MacKie, 1990] was recorded in the database at the time of acquiring the images. As was stated in section 2.2.1, there are some criticisms of this checklist as it tends to diagnose superficial spreading melanomas at the expense of the other subtypes. In addition, seborrheic keratoses tend to be over-diagnosed as melanomas due to their bizarre appearance, large size and preponderance to itching. The alternative would have been to devise a protocol whereby clinician(s) independently diagnosed all the lesions before excision biopsy on clinical grounds alone with the attendant problems of subjectivity and biases that invariably occurs. Thus, the Revised Seven-Point Checklist is perhaps the simplest and most objective measure of clinical diagnosis available and was used as a comparison for the purposes of this thesis. It is possible that specific features of this checklist may be used as predictors for melanoma in their own right. Hall (1992) found that if the maximal diameter of the lesion was seven millimetres or greater it proved to be a strong risk factor for melanoma. This is a very simple feature to measure and, as a result, is objective and reproducible. Furthermore, the presence of any one the three major criteria is considered grounds for referral to the dermatologist by the general practitioner and it would be useful to test this as a predictor for melanoma with the SIAscopy features. In addition to the checklist, other data may be added to the model to see if this results in an improvement. The obvious objective measure to add to the model would be the patient's age. The reason for this is that, like the majority of human cancers, the risk of melanoma generally increases with age [Mooi & Krausz, 1992a] according to the hypothesis of the stepwise progression of carcinogenesis that melanomas putatively follow (section 1.3.2.3) [Brodland, 1997]. However, there must be some degree of caution in adding the patient's age as a presumed risk factor for melanoma as the disease is the third most common cancer in the 20-34 year-old age group (written communication from Dept. Medical Statistics, Cancer Research Campaign, London) [CRC website].

Four specific clinical features were included in the model, namely a score greater than two in the Revised Seven-Point Checklist (termed 'suspicious'); a change in size, shape or colour in the lesion; a diameter greater than 6mm; and the age of the patient. In the first instance, each predictor was entered into the model separately using the stepwise logistic regression algorithm (forward selection method). Each predictor of the original model was allowed to compete freely with the new predictor as well as each other. If the new model that had been generated was different to the original model then the two models were compared. This was

performed by comparing the deviance values of three possible models: the deviance value of the original model ($\delta 1$); the deviance value for the new model generated by the forward

Table 5.8 Comparisons of Models with Clinical Data

Predictor Added	Predictors Included In Final Model	Predictors Excluded from Final Model	Model Deviance	Improvement in Deviance Over Original Model*	Significance
Original Model	**DM**, **AS**, **DB**, **CH**	*N/A*	196.248	*N/A*	*N/A*
Suspicious	**Suspicious**, **DM, AS, DB, CH**	None	192.469	4.479	P = 0.0340
Change in Size, Shape or Colour (CSSC)	**DM, AS, DB, CH**	CSSC	196.248	0	*N/A*
Diameter > 6mm (D>6)	**D>6, DM, AS, DB**	CH	190.373	8.563	p = 0.0034
Age	**Age, DM, AS, DB**	CH	186.986	12.011	p = 0.0005

Where:

DM = Dermal Melanin; **AS** = Asymmetry; **DB** = Blood Displacement with Erythematous Blush; **CH** = Collagen Holes; Suspicious = Score > 2 on Revised 7-Point Checklist

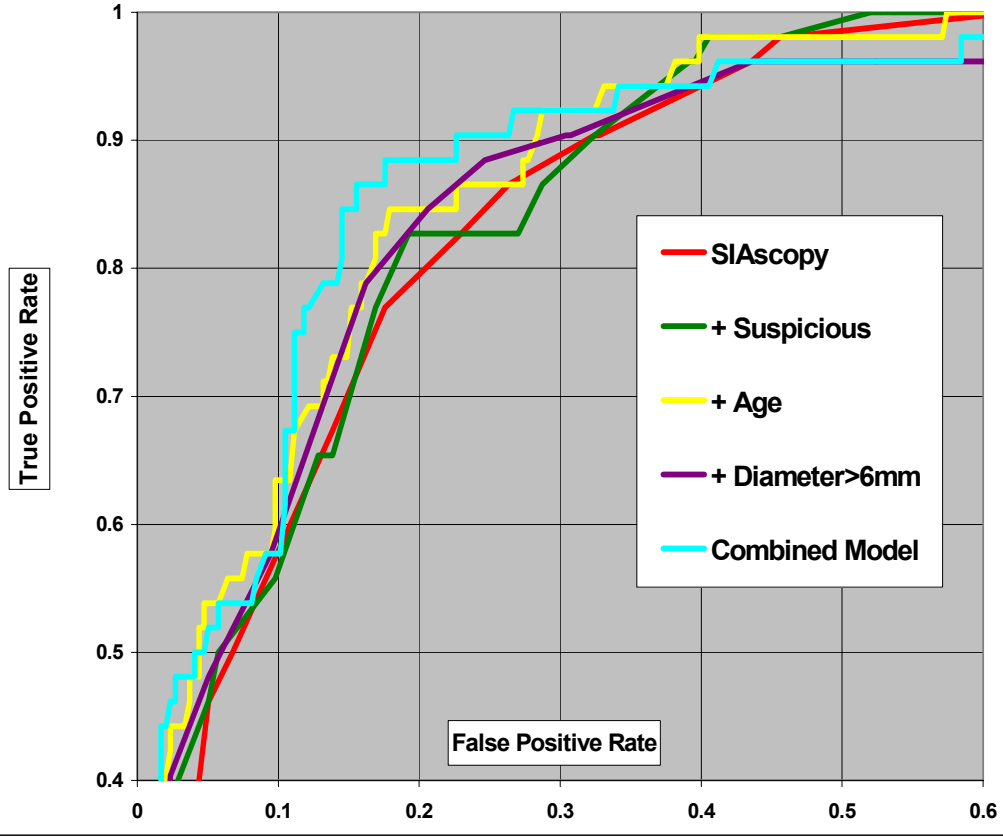
*Improvement in deviance is calculated by obtaining the deviance of the model that includes all four SIAscopy features plus the additional predictor(s). This model is constructed by using the *Forced Entry* method. This model deviance is subtracted from the model deviance of the original SIAscopy model and the difference follows a χ^2 distribution where the number of degrees of freedom is equal to the number of additional predictors.

selection algorithm ($\delta 2$); and finally the deviance value produced by generating a model with all five predictors (the original four plus the new one) by using the *forced entry* method ($\delta 3$). The model with all five predictors is a nested model of both the original model and the new model and this means that the latter two can be compared with the former. However, the new model and the original model are not nested models and so direct comparison is not possible. The difference in deviance $\delta 1$ - $\delta 3$, that follows a χ^2 distribution with one degree of freedom, represents the improvement of the new model in predicting the data with the additional predictor. The significance of this value was tested at the 5% level. Furthermore, if any predictors from the original model had not been included in the new model, the difference $\delta 3$ -

$\delta 2$ represents the worsening in prediction of the new model without the old predictor. This value follows a χ^2 distribution that can be tested for significance at the 5% level and it should *not* be significant.

With the exception of age of the patient, these were entered into the logistic regression equation as binary variables e.g. suspicious – yes or no. As before, stepwise regression models were chosen as this was exploratory work and no previous research had been performed with clinical and SIAgraphic data on which to base hypotheses for testing. It could be argued that the model developed in the previous section could be the basis for testing the

Figure 5.5: Comparisons of ROC Curves



clinical data and that the latter should be entered into the model after block entry of the four SIAscopic features. However, this method runs the risk of including features that may be no longer useful in the model when the clinical data is included. In, addition, this method would be contrary to the order in which patients are assessed clinically, namely history followed by examination and finally special investigation. The results of the forward selection logistic regression analyses are shown in table 5.8.

The addition of the clinical data produced some interesting results. Adding the predictor, 'change in size, shape or colour' did not improve the original model. This was surprising

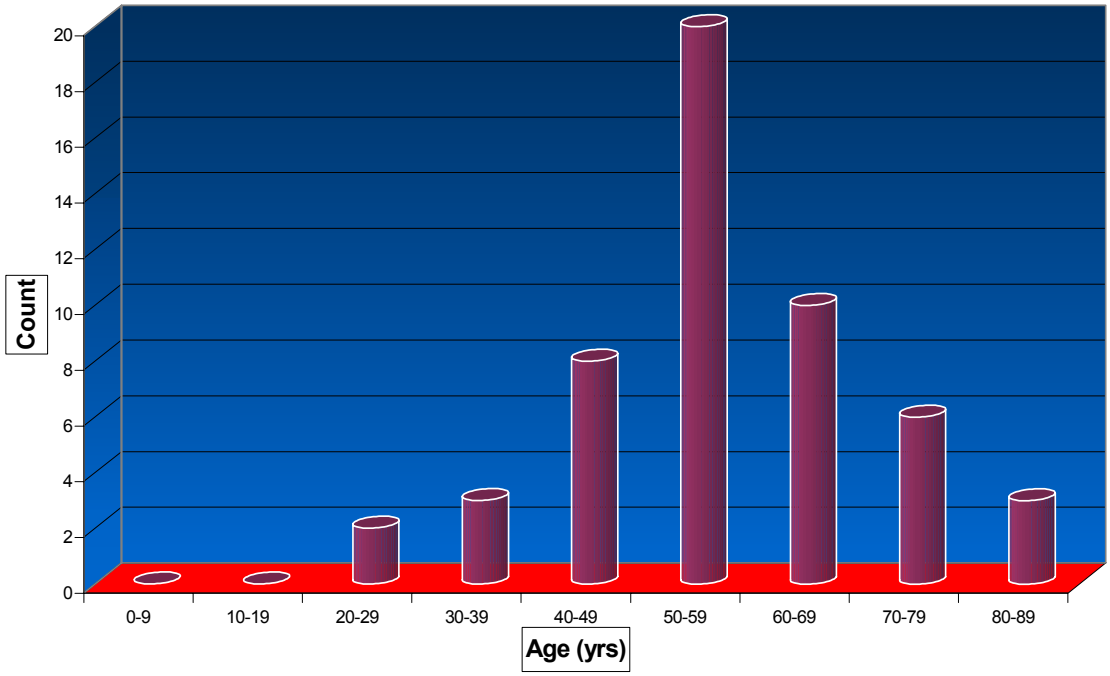
bearing in mind these three features are considered the most important in the clinical history for the revised seven-point checklist [MacKie, 1990]. One reason for this may lie in the cohort of lesions that were being investigated. The patients had been through several screenings before being entered into the trial, namely screening by the patient (they noticed something that concerned enough to make an appointment with the general practitioner), screening by the GP and screening by an experienced consultant dermatologist or plastic surgeon. It is quite likely that this stepwise screening or 'sieving' process reduced the impact of these clinical risk factors. In addition, the patients were examined from 'top to toe' by the consultant dermatologists and it was quite common for the patients to be referred for excision biopsy of a lesion other than the one that originally concerned them. As a result the patients would commonly be unaware of any change in size, shape or colour in that lesion. Lesions that were not easily visible to the patient also tended to be ignored for signs of change. In addition, some lesions changed so gradually that any recent change was not obvious to the patient and it was only through comparisons of clinical photographs that this became apparent.

For the remaining predictors, each model that was developed produced a significant change in deviance from the original four-feature, SIAscopy model. However, it must be noted that, whilst each model produced a significant reduction in deviance compared to that with purely a constant, the deviance remaining in the models was significantly different than zero. Thus, the additions of the predictors produced models that were significantly better at predicting melanoma but were still a relatively poor fit of the data (section 3.4.3.2). The addition of the 'Suspicious' predictor improved the original model, though none of the four SIAscopy features were removed. Analysis of the sensitivity-specificity pairings of the ROC curve revealed that the additional predictor increased the specificity of the model by approximately 4.5% for the same sensitivity (figure 5.5). Thus, the inclusion of the Revised 7-Point Checklist with SIAscopy served to reduce the false positive rate by weeding out the clinically benign lesions that tended to be over included by SIAscopy alone. However, according to the ROC curves, the improvement was minimal and this was borne out by the very similar curve pathway and the value of the area under the curve (0.880 versus 0.887).

Inclusion of the predictors 'Age' or 'Diameter > 6mm' also produced a significantly improved model in terms of deviance. Each of these predictors had positive regression coefficients indicating increased odds of melanoma in the presence of a diameter greater than 6mm or increasing odds of melanoma with increasing age. However, in constructing these models, collagen holes were excluded as a predictor for melanoma. There are several possibilities as to why this might be the case. It was noted in the original logistic regression model that collagen holes had the lowest odds ratio score and that the confidence intervals were close to including the value one and could be considered to be the 'weakest' of predictors out of the four SIAscopy features in the model. The theory put forward for this was that collagen holes are a feature of locally advanced disease and that half of the melanomas had a Breslow

thickness of 1mm or less indicating early local disease. It is likely that a diameter greater than 6mm and increasing age represent stronger predictors of melanoma in the early local stage of the disease than collagen holes. This is clinically and pathologically intuitive to the progression of the disease. As was stated above, with most cancers age is a strong risk factor because of the stepwise progression of carcinogenesis. Evidence that corresponds with this theory can be found in the dataset where the majority (75%) of melanomas occurred in the over fifty years old age group (figure 5.6). Early malignant melanoma spreads radially before invading vertically and an increase in diameter may be the earliest indication of a lesion in the

Figure 5.6: Distribution Of Melanomas by Age



radial growth phase [Mooi & Krausz, 1992a]. Analysis of the ROC curves for these models showed that the pathway lies above and to the left of the ROC curve for SIAscopy alone for the most part (figure 5.5). This means that these models have a higher sensitivity for a given specificity and vice versa. However, at very low specificities, the ROC curves lie below the SIAscopy curve indicating poorer performance in terms of sensitivity. This is not of concern clinically as the cut-off point for classification would almost certainly be chosen outside these regions of the curves. The areas under the curves were marginally larger ('Diameter > 6mm' = 0.885; 'Age' = 0.898) though the 95% confidence intervals included the value for the original SIAscopy model.

In the second instance, a model was generated by entering all four SIAscopy predictors and all four additional clinical predictors using the forward selection stepwise algorithm(Table 5.7). The model that was subsequently generated included the predictors dermal melanin, blood displacement with erythematous blush, diameter greater than 6mm and age. Unlike the

models that were generated previously by including a single additional predictor, it was not possible to directly compare the new model with the original. In this case, two new predictors were added and two original predictors were excluded from the new model. As a result, it is not correct to attempt to compare nested models by producing a separate one that has six predictors and analyse the difference in deviances as this does not take into account the effect of adding the predictors in one at a time, the possible interaction between the two and the effect of removing individually the original predictors. This, in effect, represents an irresolvable problem. However, given that the method used to produce the model was the forward selection algorithm using all eight possible predictors, it can be assumed that this represents the best predictive model for diagnosing melanoma using this dataset.

Table 5.9 Goodness-of-fit Statistics for 'Combined' Logistic Regression Model

Step	Variable	Step Change Deviance	Significance	Model Deviance
0	Constant	-	-	293.511
1	Blood Displacement with Erythematous Blush	49.752	< 0.001	243.759
2	Age	28.369	< 0.001	215.390
3	Dermal Melanin	22.996	< 0.001	192.394
4	Diameter>6mm	8.267	< 0.001	184.127

As expected, the predictors collagen holes and change in size, shape or colour were removed from the model. Given that the predictor 'Suspicious' only marginally improved the original SIAscopy model, it was unsurprising that this was not included in the model either. However, asymmetry was unexpectedly excluded from the final model by the logistic regression algorithm. The reasons for this are unclear but may lie in part with the fact that this feature is not as reliable and repeatable when compared to the other SIAscopy features (Tables 5.1 & 2). Furthermore, in early melanomas, asymmetry may not be a feature as apparent (or indeed present) compared to a diameter greater than 6mm – possibly the majority of melanomas have reached a diameter of 7mm before asymmetry of border and content of the lesion is apparent to the clinician on the total melanin SIAgraph. Alternatively, asymmetry may be present in sufficient numbers of benign lesions so as to reduce its usefulness as a predictor when compared to the diameter. A final explanation may lie in the close relation between a diameter greater than 6mm and asymmetry. Both features may well be indicative of the radial growth phase of the melanoma, thus the presence of one feature in the model renders the other redundant. To investigate this last point, a 2x2 contingency table

can be created to compare the predictors 'asymmetry' with 'diameter>6'. Calculation of χ^2 shows that 'asymmetry' is related to 'diameter>6' (χ^2 = 16.103, p < 0.0001). Moreover, a 2x2 contingency table constructed from only the melanomas with these predictors also shows this to be the case (Fisher's exact test: p = 0.0401).

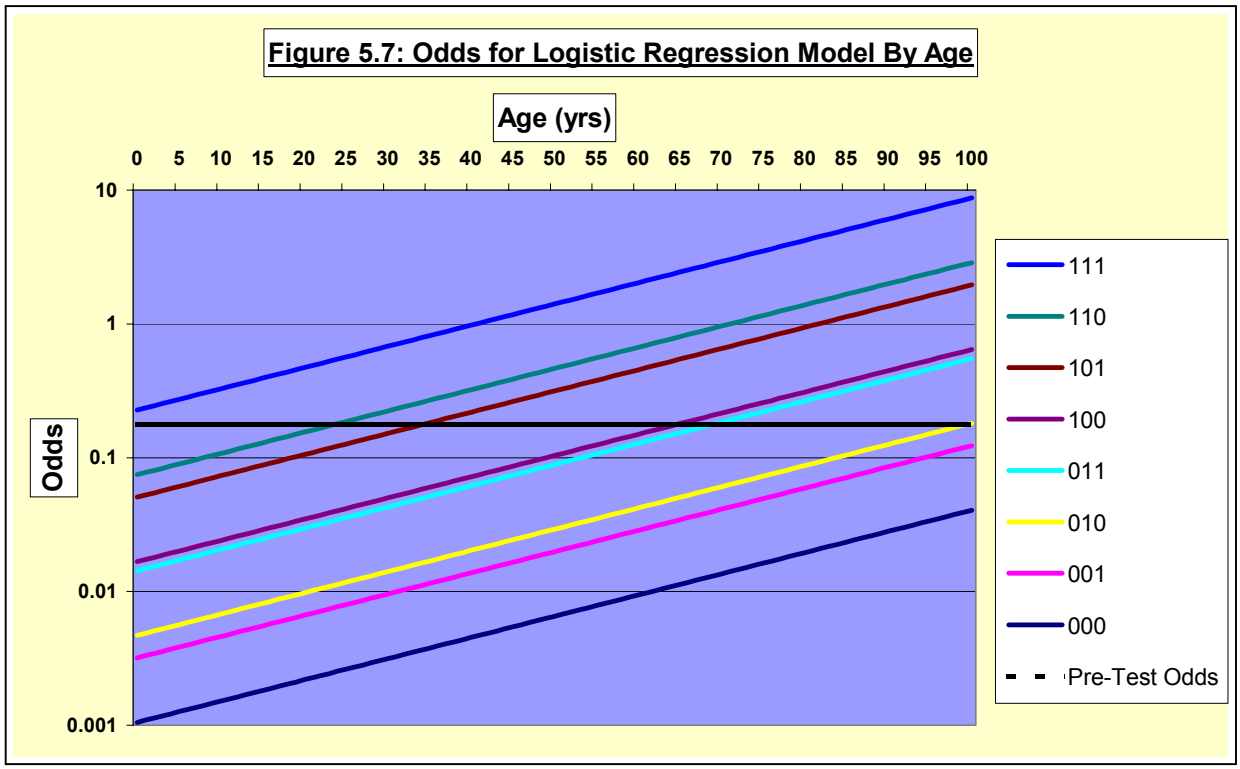
Assessment of the ROC curves showed a significant improvement in diagnostic performance as the majority of the curve lies above and to the left of the original SIAscopy model ROC curve (figure 5.5). For a given sensitivity, the specificity was nearly 10% greater in places; conversely, and more importantly, for a given specificity the sensitivity was nearly 13% greater in places. This means that at a given specificity, the rate at which melanomas were missed was reduced by 13% by this newer model compared to just SIAscopy features alone. On the other hand, at a given sensitivity (true positive rate) the number of false positives was reduced by approximately 10%. It must be noted that the curve displayed a worse performance by the newer model at lower specificities but, again, the cut-off point for diagnosing melanomas would not be chosen at low specificities. The area under the curve is 0.900 (95% confidence intervals = 0.856 – 0.944; maximal area = 0.900), which means that a randomly chosen lesion from the melanoma group will have a higher logit value than a randomly chosen lesion from the non-melanoma group 90% of the time [Campbell & Machin, 1999]. However the results must be interpreted with some degree of caution as the 95% confidence intervals include the value calculated for the original model. Thus, it is impossible to say that the difference in the area under the ROC curves did not occur by chance alone. However, given that the difference in the deviance in is highly significant, the inclination is to accept this newer model as a better predictive model for melanoma.

Table 5.10 'Combined' Logistic Regression Model Statistics

Feature	Regression Coefficient, β	Wald Statistic	Significance	Pseudo R-Value	Exp(β) (Odds Ratio)	95% CI for Exp(β)	
						Upper	Lower
Constant	-6.8567	55.2150	> .00001	-	-	-	-
Blood Displacement with Erythematous Blush	1.4979	14.6520	0.0001	0.2076	4.4724	2.0770	9.6304
Dermal Melanin	2.7654	13.1239	0.0003	0.1947	15.8848	3.5581	70.9156
Diameter > 6mm	1.1115	13.3089	0.0003	0.1422	3.0388	1.4025	6.5840
Age	0.0365	4.6973	0.0320	0.1963	1.0372	1.0171	1.0578

Collinearity analysis revealed that the Eigenvalues and condition indices are similar for each predictor and that the uncentred cross-product matrix is well balanced. These facts indicate that the model is stable and that collinearity within it is insignificant.

All of the predictors in the final model are binary except for age. Therefore, instead of



producing an odds table such as table 5.7, the information is best presented as a series of lines on a graph as in figure 5.7 (note the logarithmic scale of the y-axis). Each line is indicated by a three digit value as a shorthand way of writing the values of Dermal Melanin, Displacement with Erythematous Blush & Diameter>6mm respectively. Therefore, a value such as '011' means that Dermal Melanin = 0, Blood Displacement with Erythematous Blush = 1, Diameter> 6 = 1. The lines on the graph appear to group themselves into pairs. Examining the beta values of the predictors, it would appear that there is one predictor, namely dermal melanin that has a larger odds ratio. For purposes of discussion this will be termed the 'major binary predictor'. In addition, the other two predictors, namely diameter>6mm and blood displacement with erythematous blush have similar but lesser values for the odds ratio and these can be termed the 'minor binary predictors'. The graph indicates that patients at risk of melanoma (in the popular, non-statistical sense of the word 'risk') occur in discrete groups. At highest risk are the patients with all three binary predictors present; the next highest risk group has one major and one minor binary predictor; the next has one major or two minor binary predictors; the next has only one minor binary predictor and the lowest risk group is the one with no binary predictors at all. The line pairs have similar spacing on the graph and it would appear that 30 years of age separates each 'risk' category.

5.3.4 Choosing A Diagnostic Cut-Off

From the previous sections, two models were identified for the diagnosis of melanoma. The first used only SIAscopy features (Dermal Melanin, Collagen Holes, Asymmetry & Blood Displacement with Erythematous Blush) and the second used a combination of clinical predictors with SIAscopy features (Age, Diameter > 6mm, Dermal Melanin & Blood Displacement with Erythematous Blush). Hereafter the former model will be termed the 'SIAscopy Model' and the latter the will be termed the 'Combined Model'. It is important to assess the performance of a model containing only SIAscopy features, as this is the first clinical study investigating this technique. Thus both models were validated using the test dataset and were used in the subgroup analysis in the following section.

A diagnostic cut-off point was chosen from the ROC curve generated by the logistic regression model. The SPSS software provides a table of probabilities for each sensitivity-specificity pair like table 5.11.

Table 5.11 Probability Table for SIAscopy model ROC curve

Probability for Melanoma	Sensitivity	Specificity
0.0000	1.0000	0.0000
0.0097	1.0000	0.3547
0.0157	1.0000	0.3716
0.0179	1.0000	0.3784
0.0304	0.9808	0.5405
0.0435	0.9808	0.5439
0.0495	0.9615	0.5642
0.0550	0.9038	0.6723
0.0907	0.9038	0.6757
0.1398	0.8654	0.7365
0.1570	0.8269	0.7703
0.2361	0.7692	0.8243
0.3172	0.6731	0.8615
0.3488	0.5000	0.9324
0.4877	0.4615	0.9493
1.0000	0.0000	1.0000

The basis for deciding the cut-off point is clinical judgement and priority. Whilst subjecting patients with a benign lesion to excision biopsy that results in a scar and is often painful is undesirable, missing a curable melanoma is a much worse situation. Therefore, the diagnostic cut-off point was chosen with a very high sensitivity at the possible expense of specificity. The cut-off point for sensitivity was chosen as 95% as this was the best value obtained for skin surface microscopy in previous studies [Mayer, 1997]. This has been highlighted red in table 5.11 and the corresponding specificity is 56.42%. Table 5.12 summarises the diagnostic cut-off data for both models.

Table 5.12 Summary statistics at specific diagnostic cut-off for models

Model	Sensitivity (%)	Specificity (%)	Probability for Melanoma	Odds for Melanoma	Natural Logarithm of Odds
SIAscopy	96.15	56.42	0.0495	0.0521	-2.9550
Combined	94.23	66.00	0.0625	0.0667	-2.7081

5.3.5 Devising a Scoring System

Scoring systems for the SIAscopy and Combined models can be devised with reference to tables 5.5 & 5.10 and figure 5.8. For the SIAscopy model table 5.5, scoring values must be related to the odds ratio. Clearly the major predictor is dermal melanin. However, the product of the odds ratios when asymmetry and blood displacement with erythematous blush are present together is greater than the odds ratio of dermal melanin and this must be reflected in the scoring system that uses addition instead of multiplication. The scores were allocated thus:

SIAscopy Scoring Method

- Dermal Melanin = add **5 Points**
- Asymmetry = add **3 Points**
- Blood Displacement with Erythematous Blush = add **3 Points**
- Collagen Hole = add **1 Point**

This gives a maximum possible score of twelve. Cross-referencing table 5.11 with table 5.7 this yields a diagnostic cut-off point of greater than four points.

For the combined models it can be seen that the predictors fall into two categories – major & minor predictors and that the lines combine in pairs in figure 5.8. A method was devised that gave equal score to both Blood Displacement with Erythematous Blush and Diameter>6mm. Dermal melanin has an odds ratio that is over 3 times that for Blood Displacement with Erythematous Blush and this would be reflected in the points score. Finally, the unit change in

odds for unit increase in Age (i.e. each year) is fifteen times less than the odds ratio for Dermal Melanin. Therefore the final scoring system is as follows:

Combined Scoring Method

- Dermal Melanin = add **3 Points**
- Blood Displacement & Erythematous Blush = add **1 Point**
- Diameter > 6mm = add **1 Point**
- For every completed fifteen years of age = add **1 Point**

As very few people live to the age of 105, this gives a maximum practical score of eleven points. Cross referencing an odds table constructed from each combination of binary predictors and age in years (0-100) and a score table constructed from the same categories revealed that the threshold odds in table 5.10 consistently had a score of five or greater. This is therefore the diagnostic cut-off for melanoma. However, certain combinations of age and binary predictors also scored five yet had lower odds for melanoma. This means that the scoring system tended to over-diagnose benign lesions. Consequently, the specificity of the scoring method will be reduced when compared to that obtained from the ROC curve. This was a result of rounding errors in the calculation of the method and examination of a 2X2 contingency table for diagnosis of melanoma in the dataset revealed that there was a penalty of approximately 10% in terms of specificity. This may be deemed to be unacceptable for clinical purposes and the cut-off of greater than five points was considered instead, with a sensitivity of 90.38% and a specificity of 73.99%. These scoring methods have been reproduced in Appendix A.

5.3.6 Subgroup Analysis

Three different subsets of the original data were considered for analysis, namely 'thin' melanomas (less than 1.1 mm Breslow thickness), male and female patients. The reasons for this subset analysis were to see if there were any changes in the models especially in the early, curable stages of the disease where the clinical signs tend to be subtler. In addition, epidemiological studies have shown that there is a male:female difference in site distribution, Breslow thickness and prognosis in cutaneous melanoma [MacKie *et al.* 1997; Mooi & Krausz, 1992a]. It could be that the predictors for melanoma may change when considering the two groups separately. Table 5.13 summarises the models obtained using forward selection logistic regression analysis.

The results for the thin melanoma female subgroups were somewhat predictable. Collagen holes are likely to be a marker for locally advanced disease and this is likely to be the explanation for the absence of this predictor in the logistic regression model. In addition, Blood Displacement with Erythematous Blush was not included in the model and can be partly explained by the predictor being linked with dermal melanin. Evidence for this comes from

constructing a 2x2 crosstab table using the two predictors (χ^2 = 58.8, df = 1, p < 0.0001). Therefore, including the stronger predictor Dermal Melanin to the predictive model renders the subsequent addition of Blood Displacement with Erythematous Blush redundant. The stepwise logistic regression model generated using the female subgroup used the same predictors as the Combined model generated with all the data, albeit with different values for the regression coefficients.

Table 5.13 Subgroup Analysis Summary

Subgroup	Number of cases	Predictors in Model	Area under ROC curve
Thin Lesions	324 (28 melanomas)	Age Dermal Melanin Diameter>6mm	0.854 (0.782-0.927)
Male	124 (21 melanomas)	Asymmetry Displacement with Blush Collagen Holes	0.811 (0.707-0.915)
Female	224 (31 melanomas)	Dermal Melanin Age Diameter>6mm Displacement with Blush	0.939 (0.899-0.979)

Analysis of the male subgroup produced a predictive model based on SIAscopic features alone and, surprisingly, did not include dermal melanin. The reason Diameter>6mm was not included was likely to be because Asymmetry predicted melanoma in the same manner. A 2x2 crosstab table adds weight to this hypothesis (χ^2 = 14.23, df = 1, p = 0.00016). Similarly, a crosstab generated using Dermal Melanin and Blood Displacement with Erythematous Blush explains why the former was not included in the model (χ^2 = 20.02, df = 1, p < 0.00001). Also not included in this model was Age and this was likely to be due to the wide spread in age of the patients with melanoma (mean = 56.2yrs, standard deviation = 15.0yrs). In addition, 35% of male patients with melanoma were under 50 years of age. Finally, the model includes collagen holes. This was most likely to be because the melanomas of the male patients tended to be thicker (median Clark's level = 4, mean Breslow thickness = 1.65).

Subgroup analysis revealed that the two predictive models developed from the all of the data, Combined and SIAscopy, were mostly duplicated in the subgroups. Any difference in the models can be accounted for by subtle differences in the distribution of the predictors in those subgroups.

5.3.7 Testing the models

5.3.7.1 The Validation Dataset

A dataset for testing the predictive models generated from the investigative dataset was collected from March to June 2001. The dataset included 154 lesions (93 females : 61 males) of which thirteen were melanomas. Table 5.14 summarises the diagnostic case mix that is very similar in proportions to the investigative dataset. Of the melanomas, the majority were superficial spreading subtype and the mean Breslow thickness was 0.66mm. Furthermore, Clark's level 1 was the median indicating a mostly early local melanomas. This is the ideal test group as these are the patients that the clinician is attempting to diagnose in order to affect a cure.

Table 5.14 Diagnostic Case Mix for Validation Dataset

Diagnosis	Count
Melanoma *- Superficial Spreading* *- Nodular*	13 *12* *1*
Common Naevi (Compound, Junctional & Intradermal)	93
Dysplastic Naevi	6
Blue Naevi	1
Spitz Naevi	1
Seborrheic Keratoses	17
Lentigo	3
BCC	7
Haemangioma	3
Others	10
Total	**154**

5.3.7.2 Results of Model Validation

Table 5.15 shows the specificity and sensitivity results of the models using the validation dataset. It can be seen that both models had 100% sensitivity for the melanomas and this is

very encouraging. However, some degree of caution should be applied when interpreting these results as there were only thirteen melanomas. Consequently,

Table 5.15 Sensitivity & Specificity Analysis of Predictive Models

Model	Sensitivity (%)	95% Confidence Interval		Specificity (%)	95% Confidence Interval	
		Lower	Upper		Lower	Upper
SIAscopy	100	77.2	100	70.9	63.0	77.8
SIAscopy – Scoring Method	100	77.2	100	70.9	63.0	77.8
Combined	100	77.2	100	71.6	63.7	78.4
Combined – Scoring Method	100	77.2	100	75.9	68.2	82.2

the 95% confidence intervals for sensitivity range from 77.2% to 100%. On the other hand, six out of the thirteen melanomas were in situ (almost half of the lesions represented very early lesions) and so the dataset presents a good diagnostic challenge for the models. The specificities of both models were similar ranging from 71% to 76%. The specificity of the Combined Scoring Method (Appendix A) was slightly different form the model because a different threshold for diagnosing melanoma was chosen (section 5.3.3).

It can be concluded that the logistic regression models and the scoring systems derived from them have been validated using this dataset and it would be reasonable to suggest that these predictive models could be usefully used in clinical practice. Of the four models, the Combined Scoring Method (Appendix A) appears the best method to use in clinical practice. The reason for this is the ease with which the values of the predictors can be determined. The age of the patient is standard clinical information, the maximum diameter of the lesion can be determined by simply using a ruler and the two SIAscopy features represent the two most reliable and reproducible out of the four identified by logistic regression analysis as being useful predictors for melanoma.

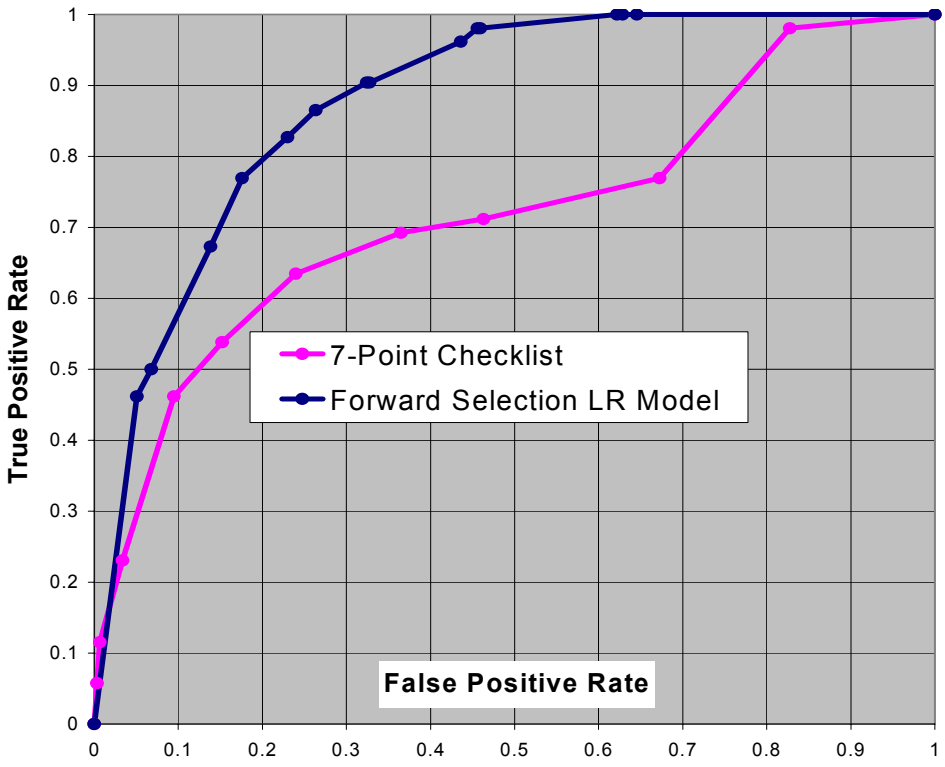
When the Combined model is considered in terms of underlying pathophysiological correlations, it can be seen that there is a predictor for each stage of the melanoma. Age represents a risk factor for change in genotype, according to the stepwise progression of

carcinogenesis model presented in section 1.3.2.3 [Brodland, 1997]. The next predictor of melanoma is Diameter>6mm. This correlates with the horizontal growth phase of the superficial spreading or acral lentiginous melanoma. Finally, the SIAscopy features are a predictor of invasive melanoma, the vertical growth phase. Thus, by combining all these predictors into one model, this predictive method represents an 'all-round' approach to the diagnosis of melanoma.

5.3.8 Comparison With Skin Surface Microscopy & Revised Seven Point Checklist

As a comparison, the ROC curve has been generated using the revised seven-point checklist [MacKie, 1990] as described in section 2.2.1. To recount, major signs (change in size, shape or colour) score two points and minor signs (inflammation, crusting or bleeding, sensory change or diameter > 6mm) score one point. This ROC curve (figure 5.8) has an area of 0.72 (95% confidence intervals: 0.635 – 0.805) and a maximal calculated area of 0.764. The area of the logistic regression model is greater than that of the revised seven-point checklist even

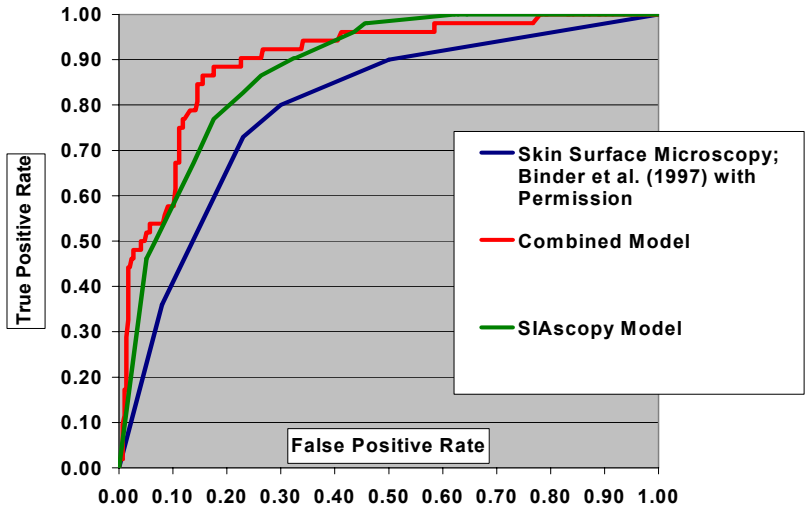
Figure 5.8: ROC Curve for Forward Selection Model



at extremes of the 95% confidence intervals and the two curves do not cross over along their length at any significant juncture, indicating that the forward selection logistic regression model performs significantly better in diagnosing melanoma on this dataset than the revised seven-point checklist.

The SIAscopy and Combined models should also be compared with skin surface microscopy. In other words, the SIAscope must be compared with the established worldwide clinical standard for diagnosing melanoma. This was not as simple as initially thought, even though skin surface microscopy images were collected from each patient. To become proficient in skin surface microscopy requires formal training of at least nine hours [Binder *et al.,* 1995] and a significant period of apprenticeship [Morton & MacKie, 1998]. Before undertaking this thesis, the author had not received formal training in dermatoscopy and any attempts to diagnose melanoma solely from these images could lead to a worse performance than diagnosing pigmented lesions on clinical grounds alone [Binder *et al,* 1995]. Therefore, it was thought that it would not be possible to appraise the skin surface microscopy images without biasing the performance of the technique. Similarly, skin surface microscopy is not universally practiced in

Figure 5.9: Comparison of Skin Surface Microscopy with Predictive Models



the UK and no clinicians at Addenbrooke's or West Norwich Hospitals felt confident enough to diagnose pigmented skin lesions from the images obtained so as not bias the comparison of the techniques. It could be argued that one of the three skin surface microscopy checklists could be employed (section 5.3.3) to increase objectivity in the of the lesions but the problem of subjectivity still remains if a lack of training means that the features in the images are not identified reliably.

Given these factors, it was decided to use indirect evidence for comparison of the two techniques. There were two sources of evidence identified: a study by Binder *et al.* (1997) and the results of the Consensus Net Meeting on Dermoscopy 2000. In the former study 11 volunteers divided into experts (previous experience and/or training in skin surface microscopy) and non-experts (clinicians with experience in diagnosing melanoma but not using skin surface microscopy). The volunteers were shown clinical and skin surface

microscopy images both before and twenty-four hours after a nine-hour training course. The dataset consisted of 100 lesions of which 33 were melanomas. The major findings were that non-experts perform worse before any training and both groups performed better after training than just diagnosing on clinical pictures alone. An ROC curve was produced that for the diagnostic performance of the groups and this is reproduced in figure 5.9. The curve that is redrawn (with permission from the publishers, Harcourt Health Sciences, USA) represents the ROC curve of the expert group after training and is the best curve on that figure. The area under the skin surface microscopy ROC curve is 0.82 (95% confidence intervals = 0.795 – 0.845; maximal area = 0.87) and its pathway is below and to the right of the ROC curves for the SIAscopy and Combined models. This would indicate that these predictive models performed better in diagnosing melanoma. In addition, the area under the curve for skin surface microscopy lies outside the bounds of the 95% confidence intervals for both predictive models adding further weight to this assertion.

In addition to the paper by Binder *et al.* (1997), results of the Consensus Net Meeting on Dermoscopy [Soyer *et al.,* 2001; Argenziano, 2001] provide further data on the diagnostic accuracy of skin surface microscopy. This involved the analysis of the four main diagnostic methods employed by clinicians, namely pattern analysis [Perhamberger *et al.*, 1987], ABCD method [Stolz *et al.,* 1994], Menzies' method [Menzies *et al.,* 1996b] & 7-point checklist [Argenziano *et al.,* 1998], in the assessment of pigmented skin lesions. It also involved the participation of 40 experts from around the world, some of whom were pioneers in this field. From this meeting it was reported that the three checklist methods showed a sensitivity of approximately 85% and a specificity of approximately 70%. For pattern analysis, the specificity was better by 12% to the next best method. The area under the ROC curves for pattern analysis, 7-point checklist and ABCD rule was reported as 0.82 and the area for Menzies' method was reported as 0.78. In addition to these data, the lack of reliability and reproducibility was quite striking. For instance, of the four criteria of the ABCD method, only Asymmetry achieved a kappa score greater than 0.4. The implications of this are put into context when considering the composition of the assessing panel.

It can be concluded that there is some evidence that predictive models based on SIAscopy outperform diagnosis based on skin surface microscopy alone, though there is the need for a structured clinical trial to formally prove this is so. What is clear though, is that formal training in skin surface microscopy takes noticeably longer and is more subjective in its interpretation than SIAscopy [Soyer *et al.,* 2001; Argenziano, 2001]. This point should be seriously considered when comparing the two methods.

5.4 Developing Predictive Models Using Classification Trees

5.4.1 CART Analysis

Classification trees using CART analysis and QUEST analysis were produced from the model-building dataset using AnswerTree (SPSS Inc., USA). Constraints set on the models were similar for both methods. The predictors entered into the models were the same as those used to develop the logistic regression models. The trees were limited to four levels, the minimum number to allow a parent node split was set to ten and the minimum child node was set to two. The priors costs were set so as to penalise the system for missing a melanoma (Section 3.6.3.1).

Table 5.16 Sensitivity-Specificity Pairs for CART Analysis

Priors		Sensitivity	Specificity	Model Complexity*
Non-melanoma	Melanoma			
0.85	0.15	0.731	0.960	Simple
0.50	0.50	0.808	0.953	Simple
0.3	0.7	0.981	0.720	Complex
0.1	0.9	1.000	0.655	Complex
0.01	0.99	1.000	0.598	Simple
*Where complexity is an arbitrary function of the number of nodes and decisions in the model				

This was done in a stepwise manner so that an ROC data for the algorithm could be built up. This method ensured that the classification tree that was clinically the most appropriate could be selected. The classification trees were generated by the CART analysis algorithm using the 'Grow and Prune' method with the '1 Standard Error' method described by Breiman *et al.* (1994). The sensitivity-specificity pairings for the varying priors setting are shown in table 5.16

It would be tempting to generate an ROC curve from these figures but, as this is not an exhaustive list of sensitivity-specificity pairs, significant error would be produced in calculating the area under the curve statistic. The best pairing is the one where the priors are set at

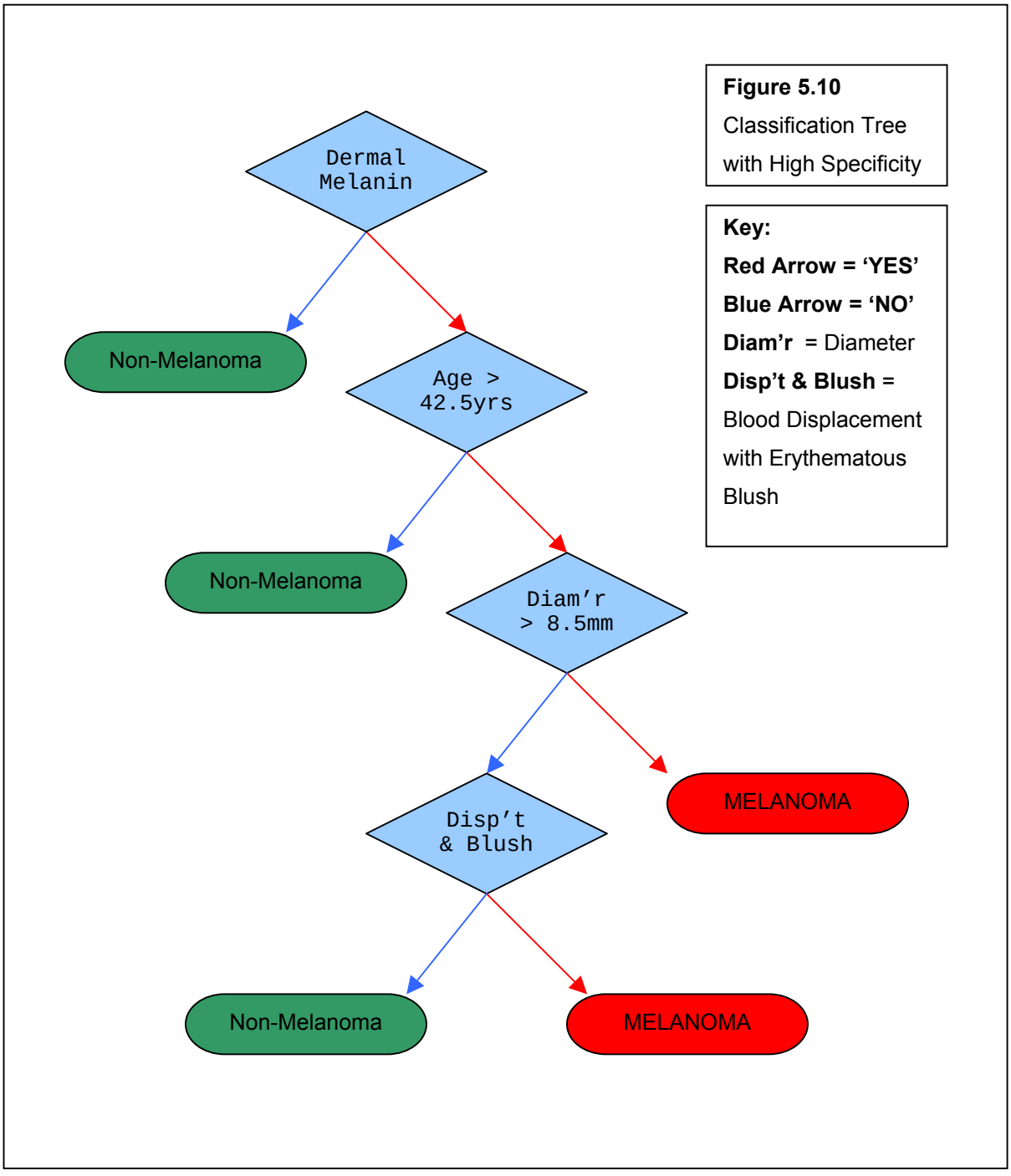
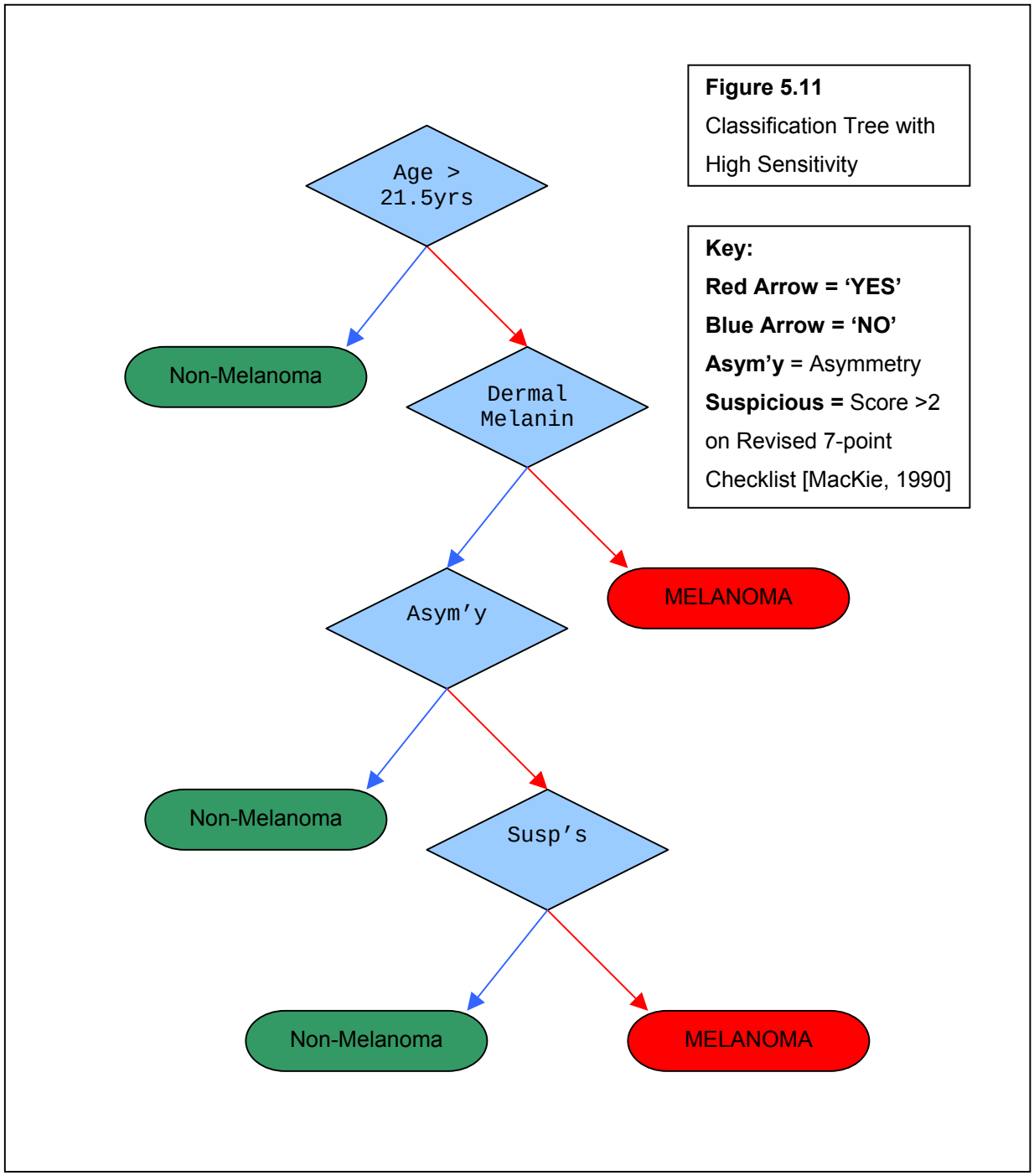


Figure 5.10
Classification Tree with High Specificity

0.3 | 0.7 for non-melanoma | melanoma respectively. However, the model generated by this set of priors is, unfortunately, too complex to be of practical use in a clinical scenario,

especially when the simplicity of the scoring method derived from the Combined logistic regression model is considered. Therefore, the two 'simple' classification trees either side of the two complex models should be offered to the clinician as predictive models for the diagnosis of melanoma. They are illustrated below in figures 5.10 & 5.11. Both classification trees require a maximum of four decisions before arriving at a diagnosis. The 'Specific' classification tree (figure 5.10) uses the same predictors as the 'Combined' logistic model regression model, though the cut-off for the diameter is different as this was entered into the



algorithm as a continuous variable. It was unsurprising that the algorithm chose the initial split on the Dermal Melanin predictor, as it had consistently been the strongest predictor in nearly

all of the logistic regression models. In addition, it was also unsurprising that the value that the Age predictor was split on was 42.5 years. As was shown in section 4.4.3, the benign naevi, namely junctional, compound & intradermal, rarely display dermal melanin after the fourth decade because the dermal melanocytes atrophy and cease to produce it [Mooi & Krausz, 1992a]. In addition, lesions greater than 8.5mm in diameter that display dermal melanin in the over 40's age group are likely to be melanomas. Moreover, those that are smaller than 8.5mm but display blood displacement with erythematous blush in that subgroup are still likely to be melanomas as this feature is a marker of invasive disease. Therefore, this classification tree would appear to be clinically intuitive and its pathway can be explained by the underlying pathophysiology of the melanoma. However, it was less sensitive than the Combined logistic regression model missing ten melanomas from the dataset. The biggest loss of sensitivity occurred at the 'Age > 42.5yrs?' decision-step where five melanomas were classified as benign. It is possible for melanomas to occur in the under 40's age group and if they are invasive they will almost certainly demonstrate dermal melanin on SIAscopy. As a result, it may be worth recommending that the clinician use the Combined logistic regression model in addition to the Specific classification tree. The sensitivity of the classification tree was, however, comparable to the results obtained from all four methods of the Consensus Net Meeting on Dermoscopy 2000 [Soyer *et al.,* 2001; Argenziano, 2001], especially when the 95% confidence intervals were taken into account (Sensitivity = 80.8%, 95% Confidence Intervals = 68.1 – 89.2%). The specificity of the tree was very high (Specificity = 95.3%, 95% Confidence intervals = 92.2% - 97.2%). Correspondingly, the false positive rate of the tree was very low and this means that the positive predictive value of the test was also high (PPV = 75.0%, Confidence intervals = 62.3 – 84.5%). Thus, for this dataset using the Specific classification tree, the probability of lesion being a melanoma given a positive test result is 75%. This value is the same as the post-test probability for the classification tree. Given this value, the post-test odds for this tree was calculated and, given the prevalence of the dataset, the Positive Test Ratio was calculated as 17.1. Any test that has a positive test ratio greater than ten is described as having a high diagnostic impact and a positive test effectively rules in the disease [Sackett *et al*., 1998]. This value was much greater than any achieved by other predictive modesl on the dataset.

Figure 5.11 outlines the decision pathway for the 'Sensitive' classification tree, so called because of its high sensitivity (100%; 95% confidence intervals = 93.1 – 100%). This sensitivity was achieved at the expense of a poor specificity (59.8%; 95% confidence intervals = 54.1 – 65.2%). The reason for this lack of specificity was the two predictors that determine the diagnosis of melanoma, namely Dermal Melanin and Suspicious, were the least specific of all, thereby generating a large number of false positives. This classification tree has a sensitivity of 100% and this would be considered ideal in the clinical situation where the physician wants to avoid the disastrous situation of missing a curable melanoma. However, the predictors used by the tree were problematic. First, the tree classified lesions on anyone

under the age of 22 as benign. Whilst it is rare, melanomas do occur in this age group. Second, the tree used asymmetry as one of the predictors. This is not the most reliable or repeatable of the SIAscopy features. Finally, the Revised 7-Point Checklist is problematic as it requires the patient to recall whether the lesion has changed in size, shape or colour. For reasons that were discussed in section 5.3.1, this information is not reliable or repeatable. Therefore, despite the very attractive 100% sensitivity of the tree, it should be recommended for use in the clinical setting with some degree of caution.

5.4.1.1 Validation of the CART models

The two CART models were validated using the same dataset as the logistic regression models. Table 5.17 shows the sensitivity & specificity achieved by the two trees.

Table 5.17 Sensitivity-Specificity Results of All Classification Trees Using Validation Dataset

Classification Tree	Sensitivity (%)	95% Confidence Interval		Specificity (%)	95% Confidence Interval	
		Lower	Upper		Lower	Upper
'Specific' CART Tree	76.9	49.7	91.8	91.5	85.7	95.1
'Sensitive' CART Tree	92.3	66.7	98.6	63.8	55.6	71.3
QUEST Tree	100.0	77.2	100.0	74.5	66.7	80.9

Both trees produced sensitivity and specificity values that were within the confidence intervals of the values obtained in the model-building dataset. It was considered that both trees had been validated by the dataset. The Specific tree, as expected, performed well in terms of specificity. However, the sensitivity was poor, though this is, in part, due to the small sample size of melanomas. This is reflected in the wide range in the 95% confidence intervals. The Sensitive tree, despite having 100% sensitivity for the model-building dataset, missed one of the thirteen melanomas in the validation dataset. This adds weight to the cautions expressed about the tree in the previous section. However, the specificity for the validation set was better than predicted, though still within the bounds of the original 95% confidence intervals. When compared to the logistic regression models (table 5.15 versus table 5.17), the CART classification trees did not perform as well in validation. Thus, so far, the Combined scoring method would still be the predictive model of choice in the diagnosis of melanoma using SIAscopy.

5.4.2 QUEST Analysis

In much the same way as the CART analysis, adjusting the prior probabilities of the algorithm generated sensitivity–specificity pairs for QUEST analysis. The results of this are displayed in table 5.18. As before, adjusting the priors to penalise the algorithm for missing melanomas increased the sensitivity at the expense of a worse specificity. Only one classification tree has

a high sensitivity with an acceptable specificity and is also simple enough to be applicable in the clinical setting. This tree is illustrated in figure 5.12.

Table 5.18 Sensitivity-Specificity Pairs for QUEST Algorithm with Varying Priors

Priors		Sensitivity	Specificity	Model Complexity*
Non-melanoma	Melanoma			
0.85	0.15	0.731	0.960	Complex
0.50	0.50	0.78.8	0.909	Simple
0.3	0.7	0.923	0.777	Simple
0.1	0.9	0.981	0.605	Complex
0.01	0.99	1.000	0.355	Simple
*Where complexity is an arbitrary function of the number of nodes and decisions in the model				

This classification tree (hereafter the 'QUEST tree') uses the same four predictors as both the Specific CART tree and the Combined logistic regression model. However, in order to increase the specificity of the tree, an extra decision node is added compared to the CART-

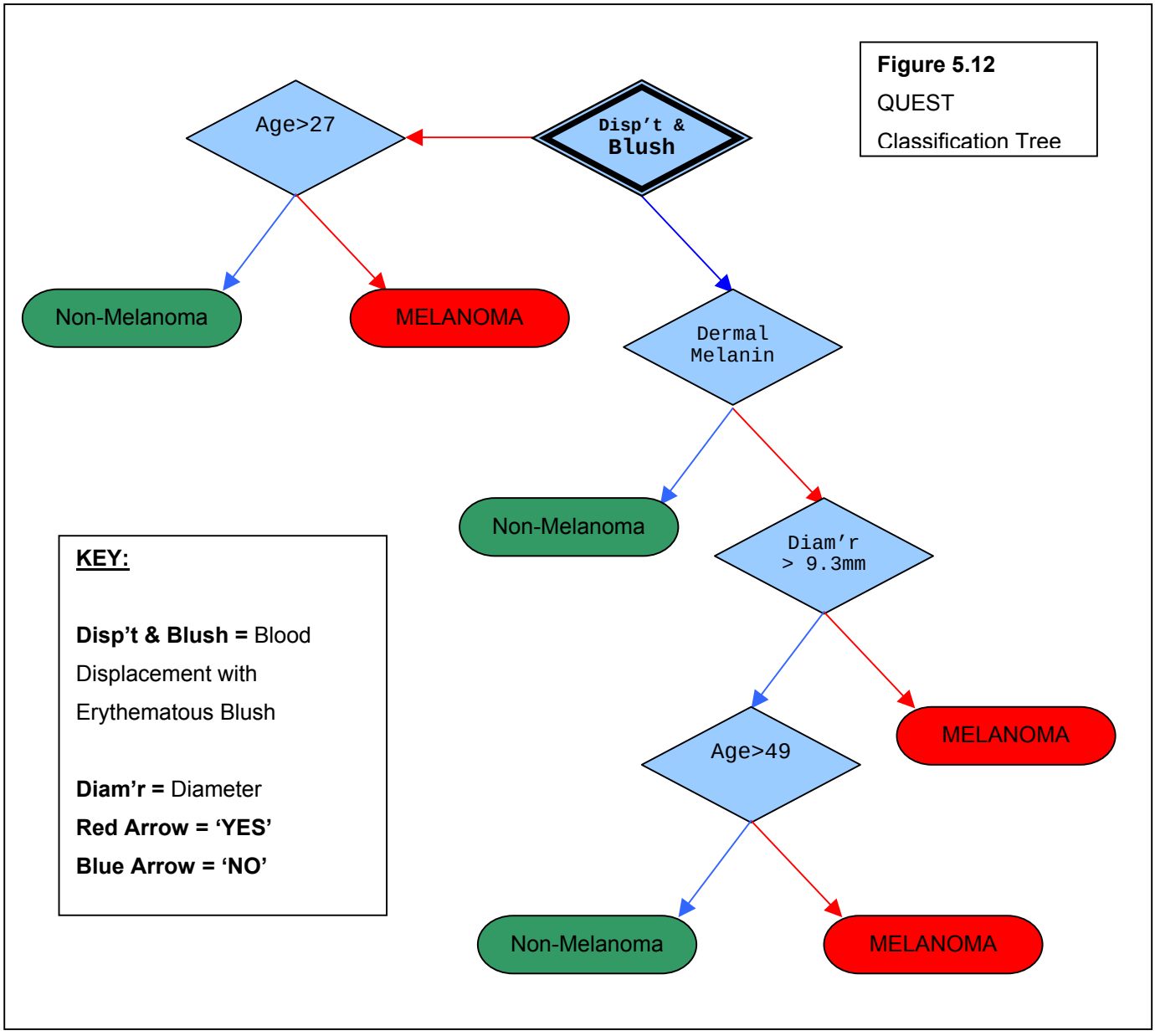


Figure 5.12
QUEST
Classification Tree

generated classification trees. As with the Specific CART model, the QUEST tree (Appendix A) is clinically intuitive in its use of predictors to classify the lesions. Interestingly, the QUEST model chooses to assess the lesion using the predictors in a different order to the Specific CART model. The QUEST tree assesses the lesion for SIAscopy signs of invasive disease and when these are found to be not present the lesion is classified as a non-melanoma. If either of the SIAscopy features is present, the lesion is assessed for the two additional risk factors, namely age and diameter. If either of these is subsequently present then the lesion is classified as a melanoma. This sifting method represents an efficient way of classifying the lesions and add weight to its clinical usefulness.

The QUEST tree (Appendix A) was validated with the validation dataset and the results are shown in table 5.17. The performance of the tree was equivalent or better than the predictions of table 5.18. Therefore, it was considered that the model had been validated. Of the three classification trees selected as candidates for putative clinical use, the QUEST tree would appear to be the model of choice. The reasons for this decision include it being the only tree to detect all the melanomas in the dataset, it having an acceptable specificity, it appearing to be an efficient way of classifying the lesions and it being clinically intuitive.

5.4.3 Comparison with Skin Surface Microscopy

Direct comparison of the performance of the classification trees was not possible by the use of ROC curves. This was because each model had only one specificity-sensitivity pair, and, unless a separate ROC curve had a point that passes through either the corresponding sensitivity or specificity, direct comparison was not possible [Zweig & Campbell, 1993]. However, indirect assessments were possible by comparing the specificity-sensitivity pairs with those of the other methods [Argenziano, 2001]. The sensitivity of the QUEST tree (Appendix A) compared favourably with all four skin surface microscopy methods. However, the lower bound of the 95% confidence interval (81.8 – 97.0%) for the QUEST tree included the values of all four methods so it was not possible to conclude that the performance is superior. It was possible to conclude, however that the QUEST tree appeared to perform at least as well as skin surface microscopy in terms of sensitivity. The specificity of the QUEST tree appeared to be superior to the three skin surface microscopy scoring methods, namely the ABCD method [Stolz *et al.,* 1994], Menzies' method [Menzies *et al.,* 1996b] & the 7-point checklist [Argenziano *et al.,* 1998]. Furthermore, the lower bound of the 95% confidence interval did not include the specificity values for these methods. However, as the 95% confidence intervals have not been quoted for the skin surface microscopy methods [Argenziano, 2001] and, given that the values were marginally lower than the lower bound of the QUEST tree, any claims of superiority must be expressed with caution. Similarly, though the value of the specificity of Pattern Analysis was above the upper bound of the QUEST tree, any claims of superiority of this method should be treated with caution. Thus, it would best to suggest that, according to indirect evidence, the QUEST tree performs at least as well as four

of the established diagnostic methods in skin surface microscopy. Once again, it is important to iterate that the QUEST tree uses simple SIAscopy predictors that are highly repeatable and reproducible in addition to standard and highly reproducible clinical information such as age and maximum diameter of the lesion. In comparison, skin surface microscopy requires a significant period of training [Binder *et al.*, 1997] to identify features that are often subtle and lack repeatability and reproducibility [Argenziano, 2001; Soyer *et al.,* 2001].

5.5 Summary

In this chapter, the main results of this study were presented. The study used two datasets, a model-building one consisting of 348 lesions and a validation one consisting of 154 lesions. Both datasets were representative of the standard caseload of a Consultant Dermatologist in a pigmented skin lesion clinic. In addition, the mix of melanomas in terms of subtype and stage (as determined by Breslow thickness and Clark's level) were sufficiently proportioned, not only to reflect the case mix in a pigmented skin lesion clinic, but also to assess the SIAscope features at identifying melanomas at different stages of local invasion and in different regions of the body.

Initially, SIAscopy features were tested for inter- and intra-observer agreement using the kappa statistic. It was found that the four main features that were identified as likely to be useful predictors for melanoma, namely Dermal Melanin, Blood Displacement with Erythematous Blush, Asymmetry and Collagen Holes, had excellent or almost perfect agreement [Kianifard, 1994]. In addition, other features were abandoned from use in further analysis as a result of their low kappa scores.

Predictive models using logistic regression analysis were generated. Originally, a model was produced using only SIAscopy predictors. This model included Dermal Melanin, Blood Displacement with Erythematous Blush, Asymmetry and Collagen Holes. Of the four predictors, the presence of dermal melanin had the largest odds ratio, the presence of collagen had the smallest and the remaining two had similar and intermediate odds ratios. This predictive model was termed the 'SIAscopy' model. Further logistic regression models were then generated using additional clinical information that would be routinely collected from the patient during their consultation. As a result, a predictive model was generated that diagnosed melanoma significantly better than the SIAscopy model. Similarly this model, termed the 'Combined' model as it combined clinical and SIAscopy predictors, used four predictors, namely Dermal Melanin, Blood Displacement with Erythematous Blush, Diameter > 6mm and Age.

The inclusion of these particular predictors appeared to make sense clinically. Of the clinical predictors both the maximum diameter of the lesion [MacKie, 1990; Hall, 1992] and the age of the patient [Mooi & Krausz, 1992; Brodland, 1997] have been shown to be risk factors for

melanoma elsewhere. Of the SIAscopy features, Asymmetry had been suggested as a marker of the radial growth phase of the melanoma; Dermal Melanin and Blood Displacement with Erythematous Blush had been suggested as markers of early invasion with proposals of the underlying pathophysiology offered; and Collagen Holes had been suggested as markers of advanced local disease. Subgroup analysis added some evidence to these suggestions. Models generated using only thin melanomas did not include Collagen Holes. Nearly all the models generated included Dermal Melanin and Blood Displacement with Erythematous Blush indicating that they were likely to be markers of early disease. Other models that were generated showed that Asymmetry and Diameter > 6mm were interchangeable and a χ^2 analysis demonstrated that these features were related. This added evidence that both features were markers of the radial growth phase of the melanoma. The Combined model was chosen as the best one for predicting melanoma. It can be thought of as using two risk factors for melanoma (Age & Diameter>6mm) in combination with two SIAscopy markers of early invasion (Dermal Melanin and Blood Displacement with Erythematous Blush).

Diagnostic cut-offs were chosen for both models that represented a compromise between over-classifying benign lesions (false positive rate) with the need to diagnose as many melanomas as possible (true positive rate). Scoring methods were devised that were derived from the respective logistic regression models and their thresholds for diagnosis were chosen accordingly. Scoring methods are popular with clinicians as they represent a quick-and-simple method of assessing a pigmented skin lesion [MacKie, 1990; Stolz *et al.,* 1994; Argenziano *et al.,* 1998; Soyer *et al.,* 2001]. Each one of these predictive models was validated using the validation dataset where it was found that the specificity and sensitivity of the models was similar to the predicted value obtained from the model-building dataset. The models had been validated. Comparisons of the predictive models were performed using the ROC curves of the methods. All of the models outperformed the Revised 7-Point Checklist [MacKie, 1990] that is an objective measure of diagnostic performance using simple clinical predictors. In addition, indirect evidence was obtained that indicated that when skin surface microscopy is used alone [Binder *et al.,* 1997] by formally trained clinicians, the Combined predictive model outperforms this technique.

Sets of classification trees were generated using CART and QUEST algorithms that were progressively penalised for missing melanomas by adjusting their ‘priors’. Out of these sets of trees, three were identified that met the acceptable specificity and sensitivity criteria as well as being sufficiently simple to apply clinically without recourse to complex flow charts. Of the two trees generated by the CART algorithm, one was highly specific for melanoma and the other was highly sensitive. On the other hand, the QUEST algorithm (Appendix A) generated a classification tree that was both highly sensitive and had an acceptable specificity. When it came to validation of the trees, the CART-generated trees performed poorly. Whilst their sensitivity and specificity values were within the bounds of the 95% confidence intervals of the

corresponding values derived from the model-building dataset, it would be difficult to recommend their use in the clinical setting. Conversely, the QUEST tree (Appendix A) performed to a standard that was equally comparable to the Combined model and scoring method (Appendix A).

Therefore, two predictive models were identified that have the advantages of being simple to learn, quick-and-easy to apply, highly sensitive and acceptably specific and at least as powerful (if not, more so) at diagnosing melanoma as established and current medical practice. These were the Combined Scoring Method and the QUEST classification tree (Appendix A).