MEGA JOURNAL OF CASE REPORTS

OPEN ACCESS

Research Article

Compiled Date: June 06, 2025

Machine Learning Classification of Skin Lesions Using Thermal Product Biosensing: A Preliminary Diagnostic Approach

Nathalie Nick[1]*, Joe Kirkup[1], Marcus Allen[1], Parv Sains[2] and Kam Chana[1]

[1]Department of Engineering Science, University of Oxford, United Kingdom

[2]Sains Surgical. 4 Elsworthy, Thames Ditton, Surrey, KT7 0YP, UK

***Corresponding author**: Nathalie Nick, Department of Engineering Science, University of Oxford, Osney Mead, Oxford, OX2 0ES, United Kingdom

Abstract

Early detection of skin cancer remains a critical challenge in global healthcare, with current diagnostic methods often suffering from delays and invasive procedures. This study explores the potential of machine learning algorithms to classify skin lesions using thermal product (TP) measurements, introducing a novel approach for rapid and potentially non-invasive skin cancer diagnosis. Leveraging data from a pilot study involving 12 patients, two primary machine learning methodologies were investigated: Logistic Regression and Support Vector Machines (SVM). The research demonstrates the potential of thermal product differences as a biomarker for skin cancer classification, with both algorithms achieving 92% accuracy in preliminary tests. The study uniquely explores both binary and multiclass classification approaches, revealing promising insights into the relationship between thermal properties and cancer characteristics. Key innovations include an exploration of logistic regression and SVM methodologies, including linear and non-linear classification techniques. The research highlights the potential of thermal product sensing as a diagnostic tool, with the ability to distinguish between different types of skin lesions based on their thermal characteristics.

Keywords: Skin cancer; Machine learning; Thermal product sensing; Diagnostic classification; Support vector machines; Logistic regression; Biomarker analysis

Introduction

As demonstrated in the companion paper Nick et al. [1], the Thermal Product (TP) Sensor offers a novel approach to quantitative tissue characterization through thermal property measurements. Building upon this technological innovation, this study explores the application of machine learning algorithms to transform raw thermal data into a potential diagnostic tool for skin cancer classification. Early skin cancer detection remains critical, with current diagnostic methods

suffering from time-consuming and invasive procedures. The TP sensor provides a unique opportunity to generate quantitative biomarker data that can be leveraged through advanced computational techniques. This research focuses on developing machine learning classification models that can interpret thermal product measurements to distinguish between cancerous and non-cancerous skin lesions. Specifically, we investigate two primary machine learning approaches –Logistic Regression and Support Vector Machines (SVM) – to assess their potential in skin cancer diagnostics. By applying these algorithms to a pilot dataset, the aim was to explore the feasibility of using thermal product differences as a diagnostic indicator and AI analysis and develop a foundation for future research.

Diagnosis

Potential classification algorithms for diagnostics were explored to evaluate their suitability for skin cancer diagnostics using the measured thermal products of patients. The objective for the algorithm output was to provide interpretable results and enable near-instantaneous diagnosis using the sensor. This would yield a diagnostic sensor that is accessible to all medical practitioners, unlike existing skin cancer diagnostic methods. The following machine learning algorithms were investigated and compared:

- Logistic Regression Model
- Support Vector Machine (SVM)

As discussed in Nick et al. **[1],** normal skin, scarring, and cancerous lesions exhibit distinguishable TP values that are measurably different from one another, indicating that TP can serve as a biomarker feature for the algorithm. However, the availability of TP data from skin lesions of actual patients represents a significant limitation, as this method of skin diagnosis is novel. Consequently, the only available data derives from a pilot study **[2]** published in 2024, comprising data from excised lesion samples obtained from 12 unique patients, as presented in **Table 1**. The data were acquired using a similar TP-based sensor, and the official clinical diagnoses for the lesions were established following TP measurement via histological examination **[2].**

Table 1: Results from 12-patient study **[2].**

Patient	Sex	Age	Histological Result	Normal Skin TP	Abnormal Skin TP	Difference
1	Female	60	Lentigo Maligna	560	361	199
2	Female	90	Nodular BCC	859	826	33
3	Male	65	Lentigo Maligna Melanoma	762	379	383
4	Male	66	Scarring port SCC Excision	289	443	-154
5	Female	90	Scarring Previous Melanoma	677	450	227
6	Male	35	Scarring Previous Melanoma	1000	1206	-206
7	Male	70	Lentigo Maligna Melanoma	687	331	356
8	Female	63	Lentigo Maligna	718	443	275
9	Female	32	Scarring Previous Melanoma	787	693	94
10	Male	31	Scarring Previous Melanoma	248	286	-38
11	Male	63	Bowenoid Actinic Keratosis	531	860	-329
12	Female	58	Malignant SCC	680	650	30

The limited dataset presents several inherent constraints that significantly impact machine learning model development and evaluation. The small sample size lacks statistical power to adequately represent the underlying patient population, thereby limiting the model's ability to capture the full spectrum of disease complexity and phenotypic variability. Additionally, the dataset exhibits class imbalance, with insufficient representation of specific cancer subtypes including Basal Cell Carcinoma (BCC) and squamous cell carcinoma (SCC), which may introduce systematic bias and compromise diagnostic accuracy. Given these data limitations, the comparative evaluation of the two algorithmic approaches will not emphasise diagnostic performance metrics within the constrained dataset presented in **Table 1**. Instead, the analysis will focus on assessing the fundamental algorithmic frameworks for their theoretical applicability to skin cancer diagnosis, scalability potential, and adaptability for enhanced performance when larger, more representative datasets become available.

Logistic Regression

Logistic regression is the first method explored. It is a statistical algorithm that predicts the probability of a binary outcome based on a set of independent prediction parameters. The binary outcomes are defined as 'non-cancerous' and 'cancerous' within the probability range $P \in [0, 1]$ respectively. For a single feature model, the score t consists of two parameters:

$$t = \beta_0 + \beta_1 x_1$$

Where β_0 is the bias term and β_1 is a feature parameter. The probability P(t) of a lesion being cancerous from TP data X is expressed in the formula:

$$P(\text{Cancer}|X) = \frac{1}{1 + e^{-\beta_0 - \beta_1 x_1}}$$

The model was trained on the first 11 training data points and tested on the final test data point $y' = \sigma(\beta T * x_{test})$ for 12 iterations and achieved an accuracy of 92% (11/12) correct diagnoses while one patient (patient number 9) resulted in a misclassification. The final β values after these iterations were [0.361 1.693] for β_0 and β_1 respectively, resulting in a decision boundary at a TP difference of 24.2, as shown in **Figure 1.**

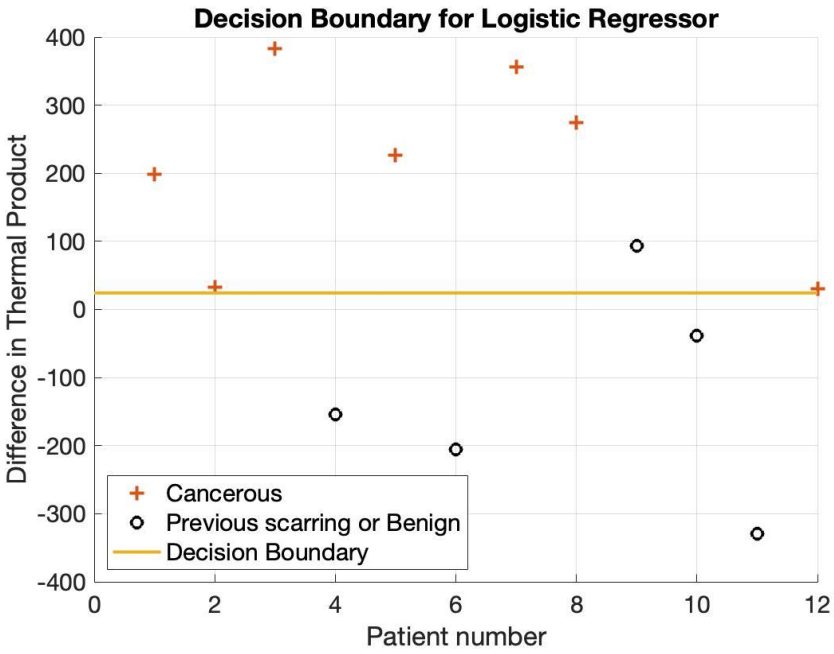

Figure 1: Decision boundary for logistic regression algorithm.

To interpret these parameters obtained, the odds and odds ratio can be used:

$$\text{Odds: } \frac{P(\text{Cancer}|X)}{1-P(\text{Cancer}|X)} = e^{(\beta_1+\beta_2 x^2)}$$

$$\text{Odds Ratio: } \frac{\text{Odds}(x+1)}{\text{Odds}(x)} = e^{\beta_1}$$

The odds ratio is commonly used in medical informatics to express the change in odds for a unit increase of the independent variable x (under normalised data, the unit increase represents one standard deviation). In this case, the odds ratio is $e^{\beta 1}$ = 5.44, which clearly shows that a higher TP difference corresponds to a large increase in the likelihood of skin cancer. The baseline odds for a patient with a lesion was also calculated to be $e^{\beta 0}$ =1.43. While the model accuracy for this small dataset was high, it may not generalise to new data as well. One reason for this is an oversimplification of the classification problem process, where only one feature, TP difference, was used as data for classification. Under this 1-dimensional problem, the decision boundary is limited to a single value and is unable to separate different classes completely.

Improved Logistic Regression

To improve the regression model, a second feature of age was added, selected due to its strong correlation with risk of skin cancer. For the two-dimensional model, the score t now consists of three parameters where β_2 was added as a second feature parameter:

$$t = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Using the same cost function, the β values were initialized as [0 0 0], updating for 3000 iterations with a learning rate of η = 0.01. The model was once again trained and tested using k-fold cross-validation, and the resulting accuracy was also 92% (11/12).

The probabilistic estimate was better than the 1-D model as the only incorrect prediction was a minor error. The final β values were [0.412 1.5234 1.2308] and an optimal decision boundary is plotted in **Figure 2**.

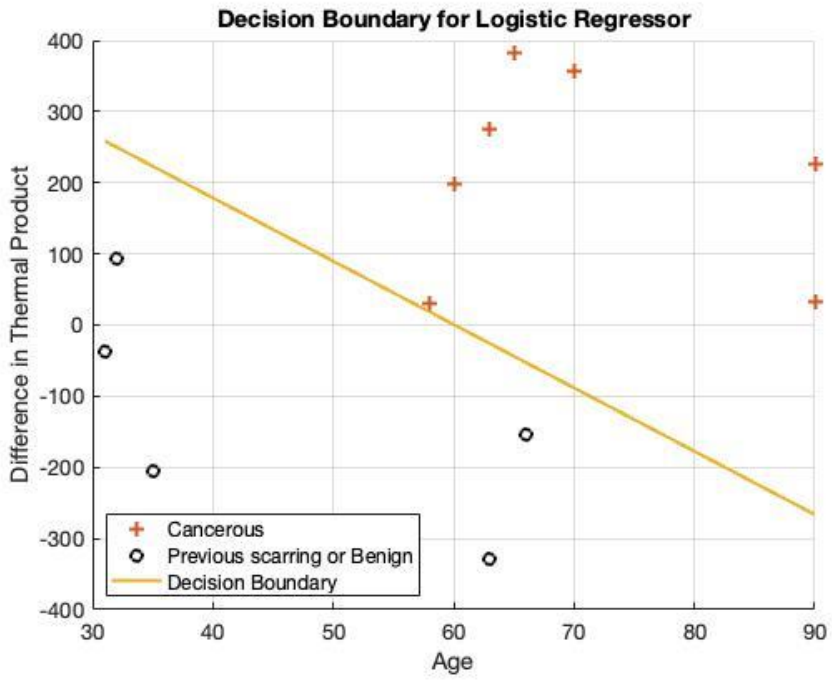

Figure 2: Decision boundary for 2-D Logistic regression algorithm.

From the decision boundary, it can be seen that the model predicts a positive correlation between age and skin cancer risk. To compare the significance of each feature, the odds ratio is used. TP difference has an odds ratio of $e^{\beta 1}$= 4.59 and age has an odds ratio of $e^{\beta 2}$= 3.42. These ratios show that while both features are indicators of skin cancer probability, the model determines TP to be a more significant feature in predicting risk. Overall, logistic regression is a useful tool in medical informatics and clinical diagnostics, and its advantages and disadvantages for the skin cancer diagnosis application using a TP sensor input are described below:

⇑**Simplicity**: Logistic regressors are computationally efficient which allows for rapid diagnosis, and is easy to implement with different features, making the algorithm adaptable with new data types.

⇑**Interpretability:** The model outputs β coefficients that indicate both the strength and direction of each feature, making it easy to understand for medical practitioners.

⇑**Probabilistic Output**: While the model can be used to directly classify data, it actually predicts the probability of skin cancer, allowing it to be used in conjunction with other diagnostic methods as part of an informed diagnosis.

⇓**Parameter Assumptions**: The model assumes that parameters are independent of each other and a linear relationship between parameters and the output. In reality, the relationship between some features (such as age [3]) and cancer risk are non-linear, and thus the model may underperform in these cases.

⇓**Sensitivity to Outliers**: The logistic regressor heavily punishes confident incorrect predictions, which can reduce model accuracy if the data consists of anomalies.

SVM Binary Classification

The SVM (Support Vector Machine) separates data into two classes by finding the optimal hyperplane that separates negative and positive data points with the largest margin possible. Unlike logistic regression, which uses a statistical approach, SVMs

are purely based on the geometric properties of the data, allowing them to generate more accurate decision boundaries. A linear SVM in binary classification was trained for classes Y = [−1, 1] corresponding to 'non-cancerous' and 'cancerous'. Hinge loss was used to maximise margins, resulting in the following objective function:

$$\min_\beta(J) = \min_\beta \sum_i^m \max\{0,1 - y_i t_i\}$$

Where J is the total cost, m is the total number of iterations, t is the score, and y is true class of the training data point (-1 or 1).

In this 2-dimensional dataset, the data is separable which allows the cost to reach 0. The resulting decision boundary and margins are plotted below with normalised axis in **Figure 3**. The support vectors, denoted by circles, are the closest positive and negative data points that lie on or within the margin lines and are used to determine the margin width. A cost of 0 is not necessarily desirable as it may be over-fitting the data, but this is an example demonstrating how accurate the decision boundary can be for an SVM model.

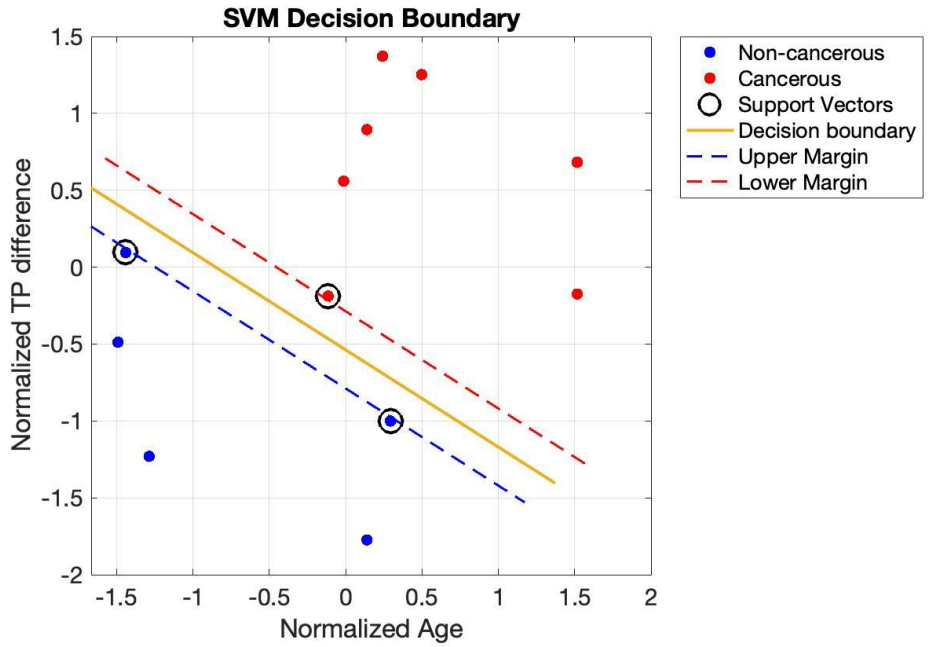

Figure 3: Hard SVM decision boundary, upper and lower margins, and support vectors.

This algorithm is a 'hard margin' SVM as it does not allow for any misclassification and produces the most accurate hyperplane possible, provided the data is linearly separable. In reality, many datasets are not linearly separable, especially as the amount of training data available increases. Therefore, a 'soft margin' should be implemented to allow some level of misclassification. This will make the algorithm more generalizable and able to handle with outliers in the data.

To do this, a regularisation term, $\lambda||\beta||^2$, is introduced to the objection function:

$$\min_\beta(J) = \min_\beta \sum_i^m \max\{0,1 - y_i t_i\} + \lambda||\beta||^2$$

Where λ is the regularisation hyperparameter. A larger λ value increases the amount of regularisation, and a value of 0 is equivalent to the 'hard margin' case. Further plots of $\lambda = 0.1$ and $\lambda = 0.5$ show the effect of varying levels of

regularisation on the margin and hyperplane **(Figure 4).**

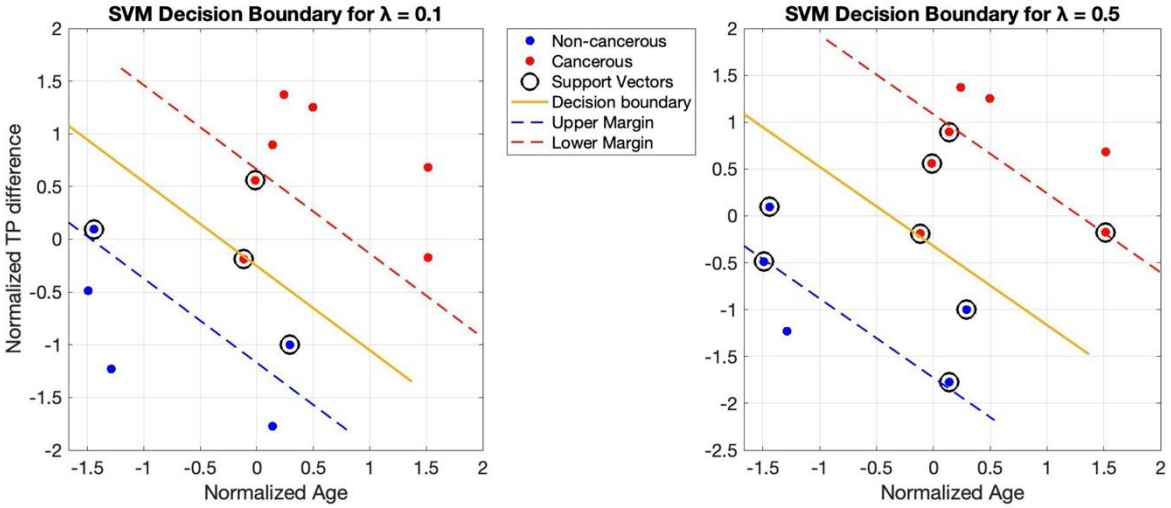

Figure 4: Soft SVM margins for varying levels of regularisation, λ = 0.1 and 0.5.

Non-linear SVM

Thus far only linear SVMs that generate linear hyperplanes have been considered. Real clinical data is often not linearly separable and can have complex non-linear relationships with each other. Therefore, to make full use of the SVM algorithm, a non-linear hyperplane can be generated using a kernel trick. The kernel trick is a technique that transforms the Euclidean feature space into a higher dimensional feature space with a kernel function $K(x_1,x_2)$, allowing the data to become more easily separable with a linear hyperplane (the resulting hyperplane would be non-linear in Euclidean space). The Gaussian or radial basis function is a common kernel function for non-linear classification that performs the following transformation:

$$K(x_1,x_2) = e^{-\frac{||x_1-x_2||^2}{2\sigma^2}}$$

Where σ is a hyperparameter. Applying this kernel trick to the binary classification problem, the resulting non-linear boundary generated is shown in **Figure 5**. While the use case for a non-linear boundary generated in **Figure 5** may not be obvious where the data is separable, the ability to produce complex hyperplanes is valuable for large clinical datasets that would not be linearly separable and have non-linear relationships. Overall, SVMs are very useful for finding accurate classification decision boundaries, and their advantages and disadvantages for the skin cancer diagnosis application are described below.

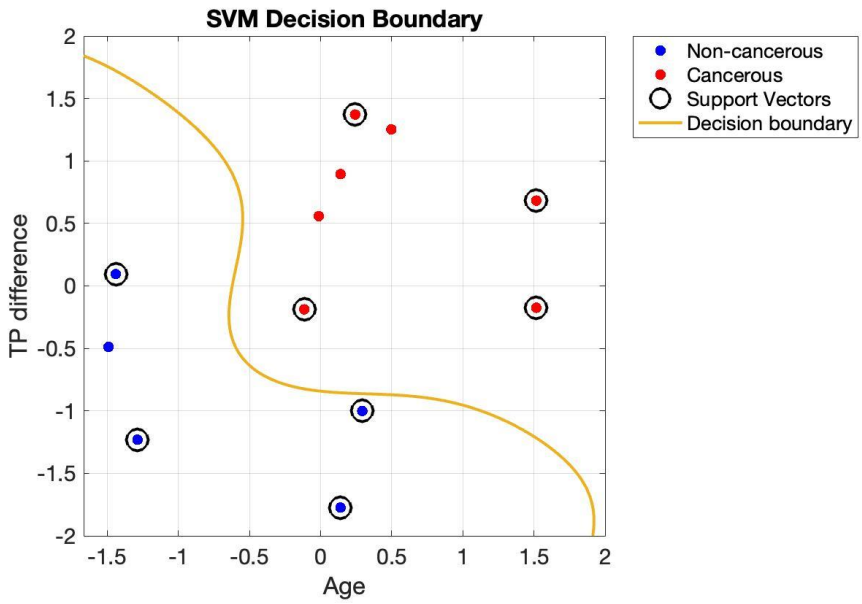

Figure 5: Hard, non-linear SVM hyperplane.

⇑**High-dimensionality**: SVMs are more effective in high-dimensional spaces, even when the number of features exceeds the number of samples. This means the algorithm can incorporate many clinical data types and be relatively reliable when the sample size is limited.

⇑**Non-linearity**: SVMs are capable of generating non-linear decision boundaries which can capture more complex relationships between features and skin cancer risk.

⇓**High Complexity**: The SVM algorithm tries to generate boundaries and margins to separate data in classes. Large datasets with high dimensionality and high noise will make this process extremely complex and reduce the model's performance.

⇓**Long training time**: Due to the SVM's high computational complexity in calculating the margins for large datasets, the training time can be extensive.

⇓**Non-probabilistic**: The hyperplane is generated based on only geometric properties of the data, meaning the SVM provides no probabilistic explanation for classification and is thus more difficult to be used in conjunction with other diagnostic methods.

Multiclass SVM Classification

Skin cancers can be classified into three types: BCC, SCC and melanoma. Melanoma is much more dangerous than the other two types, and thus it is desirable to not only distinguish between 'cancerous' and 'non-cancerous', but also specify which type of cancer a patient may have. This requires a multi-class classifier. The One-vs-Rest Method was the chosen method to perform multiclass classification with an SVM. One-vs-Rest works by training a separate binary classifier for each class, where for each given classifier, one class is considered positive and all other classes are negative, which produces a set of hyperplanes separating each class from the rest. The predicted class for a new data point would be determined by the classifier with the highest output. This technique was carried out for a 'hard' margin linear SVM. The true classes were split into three classes categorized by danger level, such that $Y \in [1, 0, -1]$ for 'melanoma', 'SCC or BCC' and 'non-cancerous' respectively. Three separate binary classifiers were then trained by implementing the

One-vs-Rest method. The hyperplane specifically classifying 'SCC or BCC' was removed as it was linearly inseparable. The resulting classification hyperplanes are shown in **Figure 6**.

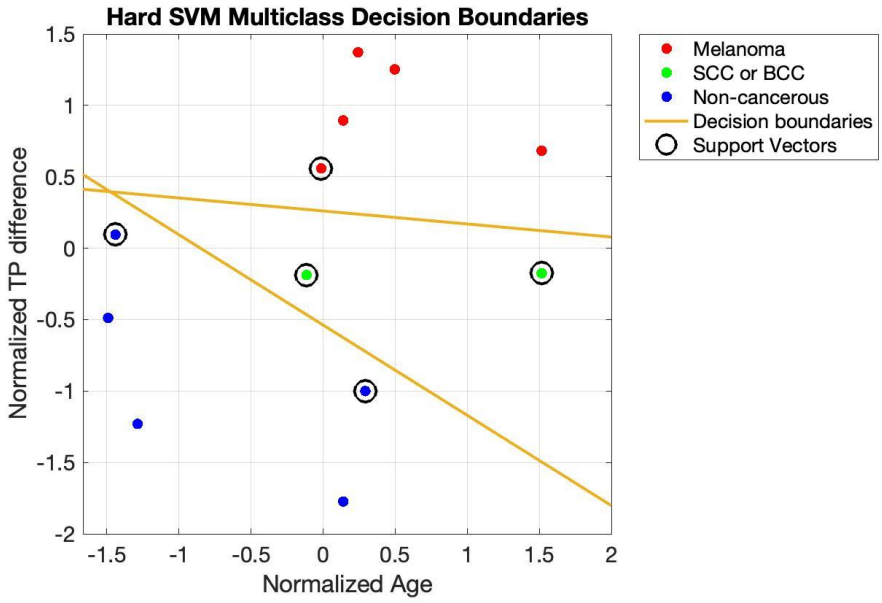

Figure 6: Multi-class linear SVM hyperplanes.

It can be observed that melanoma exhibits a larger True Positive (TP) difference compared to SCC or BCC, suggesting that higher tumor grade may be positively correlated with the TP rate. This also allows the classes to be linearly separated, which is an important discovery as it shows that TP difference can be a valid biomarker for distinguishing between different types of cancer using machine learning algorithms. This could be a significant step in developing a new diagnostic method for skin cancers.

Method

Patient data was pre-processed by first labelling for binary classification into 'cancerous' and 'non-cancerous'. Data was then normalized and split using K-fold cross-validation **(Figure 7)**.

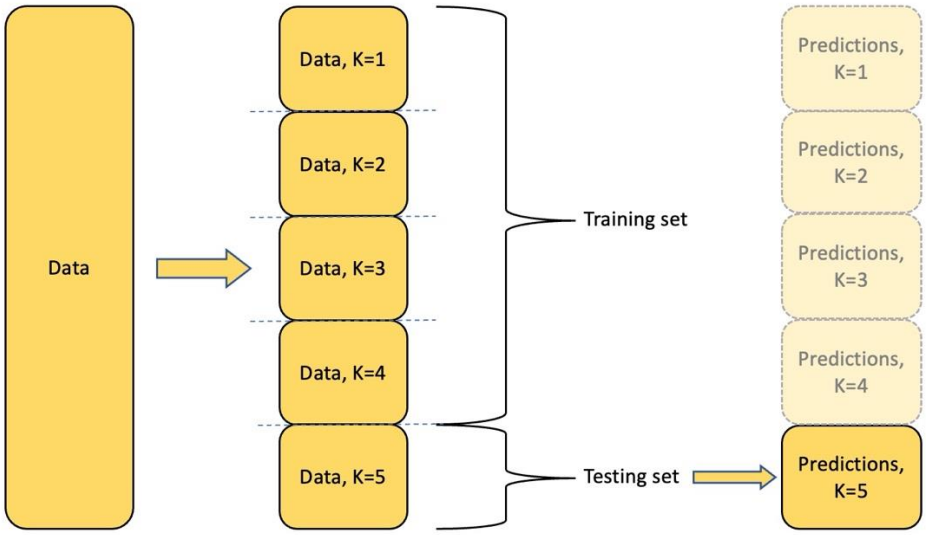

Figure 7: K-fold cross-validation illustration.

K-fold cross-validation is a training method where the data is split into k subsets. The model is trained on k-1 of the subsets and then tested on the remaining subset. This process is then repeated k times, leaving a different subset out for testing each repetition. For extremely limited data, k is set to be equal to the number of data points, i.e. k = n, which is known as leave-one-out cross-validation. While this method is computationally expensive, it maximizes the data available to produce the most accurate estimate of the model's performance and was therefore be the chosen algorithm evaluation method for this project.

Discussion

Logistic regression is well suited for early screening for skin cancer because of its simplicity, interpretability and probabilistic output, and it would serve as a useful addition to existing diagnostic methods for medical practitioners. SVMs are determined to be better suited for cases where there is a large number of features and a limited sample size, which is a common challenge in clinical medicine. Once an initial clinical trial takes place with the sensor, TP data scarcity will no longer be a limitation and quantitative analysis can then be carried out for the two algorithms. K-fold cross-validation should be used with larger subset sizes to determine the optimal hyper-parameters and model complexity. By testing different values and complexities, and then comparing their accuracy on the test subset, the optimal model can be selected. This method will also help mitigate over-fitting.

The 'age' feature for the algorithms was selected via literature review. However, modern clinical databases often contain hundreds of variables and many of these can be relevant in diagnostics, such as lesion size, skin colour (melanin content), and patient history. When more clinical data types become available, a more precise feature selection method should be used, such wrapper methods, filter methods and clinical insight from experts.

Additional machine learning algorithms should be explored for skin cancer diagnostics. As TP data availability increases, deep learning algorithms utilising neural networks should be considered, as they excel with large datasets and effectively capture complex relationships.

Conclusions

This preliminary study provides a ground-breaking exploration of machine learning techniques applied to thermal product-based skin cancer diagnosis. Despite the limitations of a small dataset, the research reveals several critical insights.

Thermal product differences demonstrate significant potential as a biomarker for skin cancer classification. Both logistic regression and SVM algorithms show promising diagnostic capabilities. The multiclass classification approach reveals potential correlations between thermal properties and cancer severity. The study's limitations, primarily the small sample size, are explicitly acknowledged. However, the results provide a crucial proof of concept for future research directions, which are: Development of more sophisticated machine learning models, collection of larger, more diverse patient datasets, and integration of additional clinical features beyond thermal product and age, and exploration of deep learning techniques as data availability increases.

The research represents an initial step towards a new transformative diagnostic approach. By combining advanced biosensing technology with machine learning algorithms, this approach offers a glimpse into future diagnostic methodologies that could provide rapid, non-invasive, and potentially more accessible skin cancer screening.

Future work should focus on expanding the patient dataset, incorporating additional clinical features, developing more complex machine learning models, and conducting clinical trials to validate the diagnostic approach.

This study underscores the potential of interdisciplinary approaches in medical diagnostics, bridging thermal engineering, biosensing, and machine learning to address critical healthcare challenges.

References

1. Nick N, et al. “Thermal Product Sensing: Simulations and Experiments of a Novel Biosensor for Quantitative Thermal Property Measurement of Biological Tissues”. Mega Journal of Case Reports. 2025.
2. DeGiovanni C, et al. Thermal Product Sensor: A potentially new diagnostic tool in the detection of skin malignancy“. Medical Research Archives. 2024.
3. P Fontanillas, B Alipanahi, N.A Furlotte, et al. Disease risk scores for skin cancers. Nat Commun. 2021;12(1):160.

Citation of this Article

Nick N, Kirkup J, Allen M and Chana K. Machine Learning Classification of Skin Lesions Using Thermal Product Biosensing: A Preliminary Diagnostic Approach. Mega J Case Rep. 2025;8(6):2001-2011.

Copyright