*Translational Data Science Journal,* 2021, pp. 1–14

**doi: 2021/TDS9**

Report

Imperial College London

REPORT

# Malignant Melanoma Prediction Using Time-Stamped Hospital Admissions in UK Biobank Participants

CIDs 01906532,[1] 01986137,[1] 02006075,[1] 01989284[1] and 01206630[1]

[1]Joint First Authors, School of Public Health, Imperial College London, Praed St, W2 1NY, Paddington, London

**Objective** To examine whether longitudinal analysis of comorbidities is able to inform time windows for prediction of malignant melanoma.

**Design, Setting and Participants** Prospective cohort study of United Kingdom Biobank enrolled participants from 2006 to 2010. A sample of 5275 incident malignant melanoma cases were matched to controls on age and sex with a 1:2 ratio by incorporating data from the National Cancer Registry. Comorbidites from Hospital Episodes Statistics admissions data were windowed using 1, 3, 5 and 10 year time periods for each case's matched controls, ensuring equal follow-up time among both groups. Melanoma risk prediction modelling was done with multivariate logistic regression, random forest, gradient boosted trees and deep learning.

**Results** Various time-windowed comorbidities marginally added predictive power to the models. The mean AUC gained from addition of the time windows across various model types was 0.021-0.027. Neural networks were the best performing model in combination with a 10-year time window, and achieved an accuracy of 0.78, precision of 0.74, recall of 0.49, F1-score of 0.59 and AUC of 0.77.

**Conclusions** Comorbidities as determined through hospital admissions are marginally informative in the short term in context of demographic and health measurements on imminent malignant melanoma diagnosis.

*Keywords:* UK Biobank, Cancer Prediction, Malignant Melanoma

## Introduction

Risk prediction modelling has been applied to various cancer types, including breast cancer, prostate cancer and pancreatic cancer. Prediction model suitability according to cancer subtype was a topic of review by Wang et al. [1]. Recent novel approaches such as incorporating phenotype risk scores and longitudinal data for prediction of pancreatic cancer have been attempted by Salvatore et al. [2]. Other approaches involved the incorporation of polygenic scores, for example in risk prediction of breast cancer [3].

Population wide data, national surveys and record linkage have enabled large data stores such as the UK Biobank (UKB) to be used for various cancer prediction models. Risk prediction modelling for various cancer subtypes, such as skin, head and neck cancers have been performed on UKB [4, 5]. Risk models are valuable in informing decisions on investigation and timing clinic appointments.

## Motivation

Melanoma is the fifth most common cancer in the UK, with an incidence of over 16,000 cases per year. A tenth of these cases are currently diagnosed at a late stage, where treatment options are limited for the patient. Detected at its earliest stage, melanoma has an excellent 5-year survival of almost 100%. Advanced malignant melanoma however, curtails a one-year median survival of just over 50% [6].

A census carried out by The Royal College of Pathologists reports that only 3% of NHS histopathology departments have enough staff to meet clinical demands [7]. Coupled with the lack of a national screening programme for the UK, melanoma may continue to be diagnosed at late stages [8].

### Prior Work

Initial melanoma risk prediction tools by Whiteman [9] sought to synthesize clinical information as predictors. Though many models have been designed, there is currently no risk prediction model for malignant melanoma in clinical use, due to the lack of sufficient validation. Olsen et al. [10], attempted validation of several models found generally high specificity (0.7-0.88) and poor sensitivity (0.37-0.64).

A recent systematic review of 40 published models over the last three decades noted usage of either traditional clinical risk factors or in combination with genetics for melanoma

risk prediction [11]. The added value of using time stamped comorbidities in addition to traditionally identified risk factors have not been examined in the context of melanoma. Here, we aimed to explore whether the addition of various time windowed comorbidities to both clinical risk factors and environmental exposures would improve predictive accuracy for malignant melanoma.

## Methods

Data cleaning and analysis were undertaken using a combination of R and Python. R version 3.6.0 and Python version 3.8.8 were used.

### Dataset

The data we received from the UKB contains behavioural information, health measurements, serum and urine biomarkers; and demographic information on over 500,000 participants. Information is collected at enrollment and at further appointments for a smaller subset of the sample population. This resource is kept up to date through linkage to National Cancer Registries (NCRAS), national death registries, and Hospital Episode Statistics (HES).

From the variables that were collected at baseline assessment, we identified groups of variables based on a priori clinical knowledge and literature on identified environmental risk factors relating to malignant melanoma [12, 13, 14].

### Study Population

We identified cases using the ICD-9 and ICD-10 codes corresponding to malignant melanoma reported through NCRAS or HES. Cases were matched to controls on a 1:2 ratio, based on age at recruitment and sex using a Mahalanobis distance metric, which takes into account the covariance structure between the matching variables [15].

### Follow-up Period for Hospital Admission

For each case, the date of diagnosis was identified and all HES events occurring on or after the date of diagnosis were removed, then time windows of either 1, 3, 5, or 10 years before diagnosis were created, with all admissions before the windows removed. The corresponding date-boundaries were applied to each cases matched control. For example, if a case was diagnosed with melanoma on 01/01/10 their windows of included admissions would be 01/01/09 - 31/12/0. 9, 01/01/07 - 31/12/09, 01/01/05 - 31-12-09, 01/01/00 - 31/12/09 for 1, 3, 5, and 10 years respectively. Then this cases matched controls would only include hes admissions from these time periods. This results in 4 datasets containing each 1, 3, 5, or 10 years of hospital admissions data per participant.

Events were aggregated according to their ICD-10 codes (Fig 1). Either to their letter, which defines a broad system category, or to letter and one number (A0) or letter and 2 numbers (A00), which correspond to more specific disease categories. Comorbidities with an occurrence of less than 0.1% in the dataset were dropped. After an initial exploration of the different follow-up times and aggregations of comorbidities, we found that the least aggregated ICD-10 (eg. A00) coding yielded better predictive power so we focused our predictive analyses on this categorisation of diseases for the 1, 3, 5, and 10 year windows.

### Descriptive Analysis

Among both groups, we looked into various confounding factors including deprivation score, smoking, alcohol, income, other cancer diagnosis and their medical history. Certain exposures were included on the basis of a priori knowledge, including sun exposure and the number of sunburn events during childhood [9]. We ran separate logistic regression models for each of these important descriptive variables, adjusting for age and gender only. Correlations between all comorbidities and the outcome of interest were calculated and visualised with a heatmap.

### Data Preparation

Data from the UKB included a series of repeated questionnaires, body measurements and biomarkers surveyed over time. Loss to follow-up meant that large proprotions of data was missing for the repeated measures. We therefore chose to exclude any features in the data-set which had over 40% missingness. Also, features that may lead to data leakage such as 'Age of First Cancer Diagnosis' and 'Number of Self Reported Cancers' were removed from the dataset.

Our case-control cohort was split so that 75% could be used for training our models and the remaining 25% for testing. All remaining missing values were imputed via simple mean imputation on the training set. Calculated means of features from the training set were then imputated within the test set. The training set was upsampled to have an even case-control split (originally 1:2). The upsampling technique chosen was Synthetic Minority Oversampling Technique (SMOTE).

### Dimensionality Reduction & Sensitivity Analysis

Modelling was done initially on the covariates identified from UKB alone, then comorbidities' time windows were added in a stepwise fashion. As part of sensitivity analysis, we also summarised the baseline variables in a lower dimensional

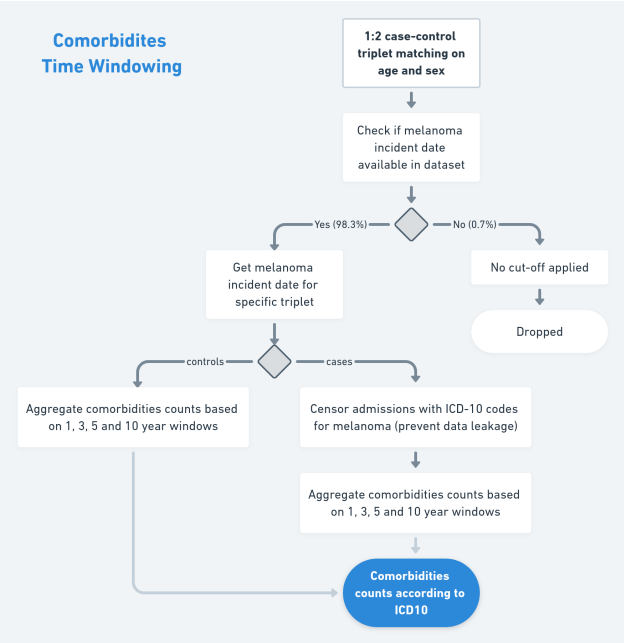



**Fig. 1.** Time windowing workflow. Admission data and dates for cases were removed before incorporation into the analysis to prevent data leakage.

form by dimensionality reduction techniques. Both principal component analysis (PCA) and factor analysis of mixed data (FAMD) were trialled on the UKB data. While, FAMD is a method of dimensionality which combines PCA and multiple correspondence analysis, we decided to focus on PCA for this study [16]. Mean imputation was performed prior to PCA. 375 principal components were chosen to represent 95% of the variance in UKB dataset. Prior to dimensionality reduction the UKB dataset had 2644 columns with continuous values.

### Predictive Models & Metrics

We took an agnostic approach to our analysis, and once our analysis datasets were ready we ran a variety of classification models and compared their accuracy, precision, recall and F1 score. We will describe the chosen models and their construction here in the methods. For all models investigated, summary information including accuracy, precision, recall, F1 score and AUC were collected, confusion matrices and ROC curves were also generated.

### Logistic Regression

To act as a baseline model, we chose a simple logistic regression model, fit with all comorbidities over described time windows, with no penalisation.

### Random Forest

Tree based classifiers were chosen because we expected a complex and non-linear relationship of the predictors. From this group, we first selected the Random Forest algorithm as a widely known and well-understood method. As the aim of this study was simply to understand what level of prediction can be achieved from comorbidity data – as opposed to aiming to create a diagnosis rubric for clinicians – we believed that an ensemble method was valid and appropriate. A variety of tree depths were tested to ascertain to optimal depth for minimal overfitting. Following this, the number of estimators, minimum sample split, and impurity metric (gini or entropy) were tuned using a grid search method.

### XGBoost - eXtreme Gradient Boosting

XGBoost implements stochastic and regularised gradient boosting in random forests to improve efficiency and performance. Default hyperparameters were initially used for modelling on the UKB feature set as outlined in the data preparation section. The models were then used to compute predictions for the 1, 3, 5 and 10 year phenotype time windows. Sensitivity analyses using PCA feature sets were done. The models using the UKB feature set without dimensionality reduction were selected for hyperparameter tuning due to better predictive performance.

The hyperparameters considered for grid-search tuning were max-depth, min-child-weight, subsample, learning-rate and gamma. During the tuning process we aimed to avoid over-fitting to our training data and minimise the false negative rate of our predictions. This was monitored by comparing training and test set AUC predictive performance.

### Neural Networks & Architecture Search

This additional approach was trialled to see if using a less interpretable deep learning method would yield a better prediction than tree-based methods.

Supervised deep learning binary classifiers were built and a neural network architecture search was performed to optimize for predictive accuracy. The number of hidden layers was varied between one and nine, with all interior neurons using the ReLU activation function. The hidden layers consisted of combinations of fully connected layers, dropout layers, and batch normalisation layers.

Different learning rate optimisers were used, including stochastic gradient descent, Root Mean Square Propogation (RMSProp) and Adaptive Moment Estimation (Adam). The complexity of the neural networks was also varied by adjusting the number of neurons contained in each layer from three to 512. Finally, L2-regularization was applied to a subset of the models. All models were trained to minimize for binary cross-entropy loss.

The input vector consisted of 1570 entries containing demographic, biomarker and comorbidity information (1285 features from HES and 285 features from UKB). The output layer for all models was a single neuron using a sigmoid activation function to classify cases and controls. All models were trained for a maximum of 2000 epochs, though an early-stopping function was employed to stop training should the loss function of the validation set exceed a threshold of 0.01 of the prior 15 epochs. The best performing network architecture on the 10-year windowed data was then used to perform prediction with various time-windowed comorbidities.

**Table 1.** Descriptive statistics on the study population.

| Feature | Cases n = 5275 | Controls n = 10550 | Total N = 15825 |
|---|---|---|---|
| **Mean Age (sd)** | 59 (7.54) | 59 (7.54) | 59 (7.54) |
| **Sex (%)** | | | |
| Male | 2417 (45.8) | 4834 (45.8) | 7251 (45.8) |
| Female | 2858 (54.2) | 5716 (54.2) | 8574 (54.2) |
| **Ethnicity (%)** | | | |
| White | 4947 (93.8) | 9346 (88.6) | 14293 (90.3) |
| Other | 325 (6.2) | 1186 (11.2) | 1511 (9.5) |
| **Mean Deprivation Score (sd)** | -1.91 (2.69) | -1.33 (3.09) | -1.52 (2.97) |
| **Mean household income pre tax, £ (%)** | | | |
| < 18k | 941 (17.8) | 2339 (22.2) | 3280 (20.7) |
| 18 – 31k | 1182 (22.4) | 2380 (22.6) | 3562 (22.5) |
| 31 – 52k | 1168 (22.1) | 2198 (20.8) | 3366 (21.3) |
| 52 – 100k | 913 (17.3) | 1483 (14.1) | 2396 (15.1) |
| > 100k | 247 (4.7) | 425 (4.0) | 672 (4.2) |
| **Skin colour (%)** | | | |
| Black | 0 (0.0) | 75 (0.7) | 75 (0.7) |
| Brown | 21 (0.4) | 266 (2.5) | 287 (1.8) |
| Dark olive | 35 (0.7) | 167 (1.6) | 202 (1.3) |
| Light olive | 577 (10.9) | 1900 (18.0) | 2477 (15.7) |
| Fair | 3904 (74.0) | 7187 (68.1) | 11091 (70.1) |
| Very fair | 680 (12.9) | 796 (7.5) | 1476 (9.3) |
| **Mean childhood sunburns (sd)** | 2.56 (4.74) | 1.42 (3.07) | 1.8 (3.71) |
| **Smoking status (%)** | | | |
| Current | 367 (7.0) | 1098 (10.4) | 1465 (9.3) |
| Never | 3012 (57.1) | 5585 (52.9) | 8597 (54.3) |
| Previous | 1870 (35.5) | 3795 (36.0) | 5665 (35.8) |
| **Mean cancer occurrences (sd)** | 3.61 (2.16) | 2.98 (1.45) | 3.19 (1.98) |
| **Deaths since enrollment (%)** | 273 (5.2) | 396 (3.8) | 669 (4.2) |

## Results

### Study Population

A total of 5275 melanoma cases satisfied our initial inclusion criteria and our matching process yielded 10,550 controls. Descriptive statistics were generated for this population as per Table 1. For the remaining analysis, 156 cases were removed, with their respective controls, due to missing values for their date of melanoma diagnosis. One further matched control was excluded due to miss-classification as a control. Our total sample size for the remaining analysis was 15356.

### Descriptive Analysis

Demographic and baseline health information is tabulated in table 1. Mean age and distribution of sex between groups are similar as expected from the matching process. Ethnicity was grouped into "white" and "other" due to low frequencies of other ethnic groups across the entire UKB. The deprivation score represents a compilation of several indicators of relative affluence in an attempt to provide an overall score for an individual's deprivation. A negative score indicates a relative worse state of deprivation. The UKB population is more affluent than the general UK population [17] and the case sample in our data is more affluent than the control sample according to its corresponding mean deprivation score. Mean childhood sunburn events were higher among cases, however this may be due to recall bias associated with cases knowing their melanoma status. Other notable proportional differences in table 1 are further highlighted in table 2.

Separate regression analyses agreed with prior research findings in regards to skin colour and melanoma risk [14]. As we might expect, non-white ethnicity and darker skin colours were associated with decreased odds of melanoma. There were no cases from the UKB with black skin, thus this odds ratio is uninterpretable. In keeping with Hertog et al. [18], current and previous smokers had decreased odds of melanoma. Individuals who drank alcohol more frequently than average had an increased odds of melanoma. Interestingly, an increased deprivation score (less affluent) was significantly associated with decreased odds of melanoma (OR 0.93, 95% CI [0.92, 0.94]). Similarly, a high salary was also associated with increased odds of developing melanoma (£52-100k 1.29 [1.15, 1.44], £100k+ 1.21 [1.02, 1.45] (OR [95%CI])). Comorbidities that were recorded at assessment were also tabulated for our study population. All events with $> 1\%$ prevalence can be found in table 4 in the Appendix.

### Correlation Matrix of Comorbidities

For due diligence, a correlation matrix was constructed before running the predictive models to ensure that no significant and unexpected correlations between the malignant melanoma ICD-10 code and other ICD-10 codes exist (Fig 2). Detection of any correlation structure here would identify any potential labelling errors, where other codes were substituted for malignant melanoma or potential duplication in admissions coding.

### Predictive models

All predictive models were developed with the UKB feature set first, then with the 1, 3, 5 and 10-year time-windowed phenotype datasets in addition (5 analysis datasets in total).

**Table 2.** Separate logistic regression analyses adjusted for age and gender only.

| Predictor | Odds Ratio [95% CI] | p-value |
|---|---|---|
| **Deprivation Score** | 0.933 [0.922, 0.944] | <0.001 |
| **Average Income** | | |
| 18,000 to 30,999 | 1 | |
| 31,000 to 51,999 | 1.092 [0.987, 1.207] | 0.087 |
| 52,000 to 100,000 | 1.287 [1.151, 1.438] | <0.001 |
| Greater than 100,000 | 1.218 [1.022, 1.448] | 0.027 |
| **Ethnicity** | | |
| White | 1 | |
| Other | 0.419 [0.354, 0.493] | <0.001 |
| **Smoking status** | | |
| Never | 1 | |
| Current | 0.618 [0.544, 0.7] | <0.001 |
| Previous | 0.912 [0.849, 0.98] | 0.012 |
| **Alcohol Consumption** | | |
| Once or twice a week | 1 | |
| Daily or almost daily | 1.109 [1.006, 1.223] | 0.037 |
| Never | 0.683 [0.59, 0.788] | <0.001 |
| 1-3 times a month | 0.984 [0.872, 1.11] | 0.794 |
| Special occasions | 0.834 [0.738, 0.941] | 0.003 |
| 3-4 times a week | 1.1 [1.001, 1.208] | 0.047 |
| **Skin Colour** | | |
| Fair | 1 | |
| Black* | < 0.001 [> 0.001, > 0.001] | 0.891 |
| Brown | 0.144 [0.09, 0.22] | <0.001 |
| Dark olive | 0.383 [0.261, 0.545] | <0.001 |
| Light olive | 0.558 [0.504, 0.617] | <0.001 |
| Very fair | 1.572 [1.409, 1.755] | <0.001 |

* 0 vs 75 participants.

### Logistic Regression

This model was used as a baseline to explore the linearity of the underlying data. The model shows the poorest performance in terms of AUC. We also note that no improvement in AUC is obtained by adding different time windows (Table 3). This implies that no additive linear predictive information for the log-odds of melanoma is held within the time windows extending up to a period of 10 years before diagnosis.

### Random Forest

In order to model non-linear relationships between our predictors and outcome we chose to use a random forest model on the data. This approach led to a mean improvement of 0.119 in AUC across the time windows. This supports the hypothesis of non-linear relationships between predictors and melanoma diagnosis. Including time-windowed phenotypical data led to a mean improvement of 0.011 in AUC. There was no meaningful improvement when increasing the time-window from 1 to 10 years suggesting that any additional predictive information provided by the phenotypical data was present within the 1 year time window.

### XGBoost

Several XGBoost models were explored to further improve our predictive accuracy. The performance reported in table 3 was after tuning the following hyperparameters: learning rate, maximum tree depth, minimum child weight and subsample rate. We see that the XGBoost models, using the full UKB feature set, performs better in terms of AUC across all

**Table 3.** Summary of scores for all predictive models with corresponding time windows.

| Model | Time-window | Accuracy | Precision | Recall | F1-Score | AUC |
|---|---|---|---|---|---|---|
| **Logistic Regression (UKB)** | NA | 0.665 | 0.354 | 0.044 | 0.078 | 0.528 |
| | 1 year | 0.665 | 0.354 | 0.044 | 0.078 | 0.528 |
| | 3 years | 0.665 | 0.354 | 0.044 | 0.078 | 0.528 |
| | 5 years | 0.665 | 0.354 | 0.044 | 0.078 | 0.528 |
| | 10 years | 0.665 | 0.354 | 0.044 | 0.078 | 0.528 |
| **Random Forest (UKB)** | NA | 0.682 | 0.600 | 0.080 | 0.141 | 0.636 |
| | 1 year | 0.679 | 0.517 | 0.122 | 0.200 | 0.652 |
| | 3 years | 0.686 | 0.563 | 0.130 | 0.211 | 0.646 |
| | 5 years | 0.680 | 0.585 | 0.043 | 0.080 | 0.636 |
| | 10 years | 0.676 | 0.500 | 0.100 | 0.167 | 0.653 |
| **XGBoost (UKB)*** | NA | 0.684 | 0.556 | 0.116 | 0.192 | 0.658 |
| | 1 year | 0.694 | 0.601 | 0.155 | 0.246 | 0.683 |
| | 3 years | 0.692 | 0.588 | 0.162 | 0.254 | 0.686 |
| | 5 years | 0.691 | 0.592 | 0.148 | 0.237 | 0.688 |
| | 10 years | 0.689 | 0.575 | 0.150 | 0.238 | 0.683 |
| **Neural Network (UKB)** | NA | 0.768 | 0.743 | 0.433 | 0.547 | 0.747 |
| | 1 year | 0.775 | 0.740 | 0.472 | 0.577 | 0.766 |
| | 3 years | 0.777 | 0.738 | 0.482 | 0.583 | 0.767 |
| | 5 years | 0.778 | 0.749 | 0.472 | 0.579 | 0.768 |
| | 10 years | 0.777 | 0.740 | 0.481 | 0.583 | 0.771 |

NA denotes that no time windowed comorbidities were used, and prediction was performed solely on covariates from UKB. *Sensitivity analysis with XGBoost on 375 principal components of the UKB covariates showed overall reduced AUC scores (NA 0.619, 1 year 0.643, 3 years 0.642, 5 years 0.644, 10 years 0.636).

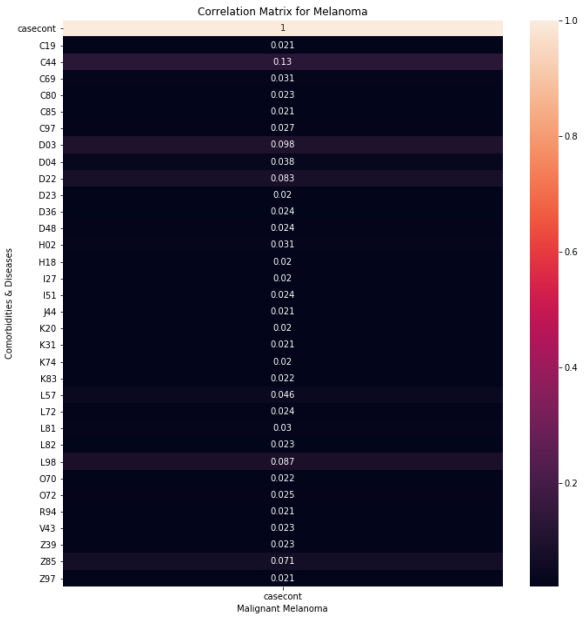



**Fig. 2.** Correlation matrix of malignant melanoma ICD-10 code "C43" against all other ICD-10 codes. No significant correlation was detected.

time windows than any of our previous models. After initial hyperparameter tuning with overall accuracy as the evaluation metric, we also aimed to decrease false negative predictions due to the clinical implications of misclassifying a melanoma case as healthy. Recall score was monitored and further hyperparameter tuning achieved an increase of 0.04 without detriment to overall accuracy. A mean increase of 0.027 in AUC was observed when adding phenotypical data for the various time windows. No significant difference in AUC was seen between the time windowed models. This reflects the findings in our previous models, but to a smaller magnitude.

We also conducted a further analysis using 375 principal components that represented 95% of the explained variance of the UKB feature set. Approaching the UKB feature set in this manner led to a mean decrease of 0.043 in AUC across all models. Similarly, to the full feature set model we observed a mean AUC improvement of 0.022 when incorporating time windowed phenotypical data in our predictive model. Overall, we did not observe any meaningful difference in performance after the 1 year time window.

In order to better understand the differences in predictive performance between the time windowing feature sets, feature importance plots were generated for each XGBoost model (appendix III). Of note, the co-morbidity data (HES time windows) has a relatively weak feature importance as seen by the sparsity in selection after feature 375 (appendix III). This becomes more apparent as we increment the time windows. The most important features for XGBoost model with 1 year time-window were childhood sunburn occasions, skin colour, ease of skin tanning and prevalence of other skin disorders (ICD-10 code L98).

### Neural Networks and Architecture Search

A manual neural network architecture search was performed, varying the networks' complexity and total parameter count from 30,000 to 3,000,000 (Fig 4). The best performing neural network was constructed with input batch normalization and one hidden dense layer containing 32 neurons. L2-regularization was applied to all layers' weights with a lambda of 0.05. A learning rate of 0.001 was used with the stochastic gradient descent optimizer. The model achieved convergence after 51 epochs (Fig 5 and Fig 6a in Appendix).

The above described architecture was similarly applied on to model malignant melanoma on the UKB dataset alone, followed by the addition of 1, 3, 5 and 10-year time windowed datasets. There was marginal predictive advantage gained from the use of different time windows, with a mean AUC improvement of 0.021 (Table 3, Fig 6b in Appendix). Neural networks predictive performance slightly improved with larger time windowed phenotypes. The best results were obtained on the 10-year windowed test set: accuracy of 0.78, precision of 0.74, recall of 0.49, F1 score of 0.59 and AUC of 0.77. Sensitivity analysis with use of PCA instead of the UKB predictors found a mean decrease in AUC for neural networks by 0.05, comparable with the findings from the XGBoost model.

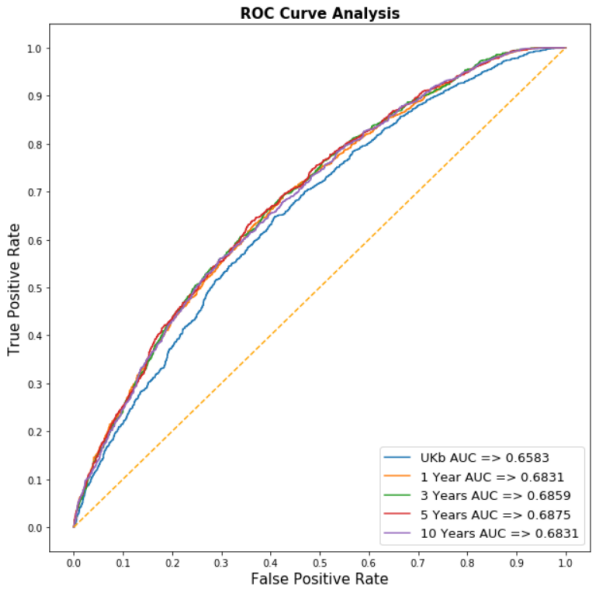



**Fig. 3.** ROC Curves with their corresponding AUC scores for our best performing tree-based model.

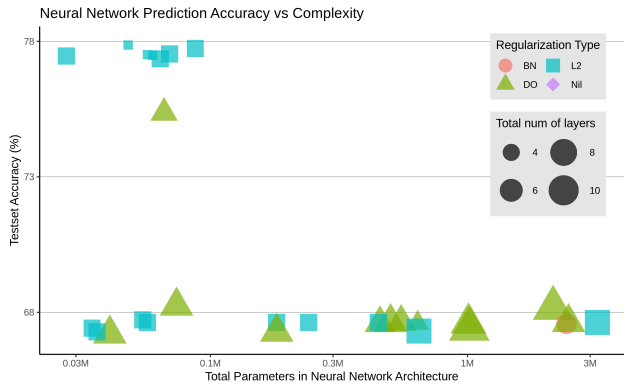



**Fig. 4.** Testset accuracy achieved by neural networks of various architecture types, with total number of layers depicted by size. Regularization approaches are color coded and denoted by shapes. BN = Batch Normalization, L2 = L2-weights regularization, DO = Dropout layers.

## Discussion

We were able to successfully build a model that predicts melanoma with a high degree of accuracy and low false negative rate. We found that tree-based models out-performed standard linear models and the neural networks outperformed both of these, suggesting a non-linear structure relationship between the phenotypic data and malignant melanoma. Our approach of using time-stamped data is novel for the prediction of melanoma, and our study showed the benefits of this approach in the short term.

With regard to our primary aim, we found that time-stamped comorbidities can improve prediction of melanoma slightly in the short term. With regard to our secondary aim, we found that in the tree-based models relied mainly on baseline variables for prediction, with a few skin related variables found to be of importance. We were able to examine the variable importance of the tree-based models, but further inspection is required to gain a deep understanding of how the neural network is outperforming the tree-based models.

A surprising result from the use of hospital admissions data was that including more years of follow up time did not significantly improve the predictive power of our models. We expected that including information on comorbidities would improve the prediction of malignant melanoma, as it had been seen in studies for predicting other cancers [2, 3]. There are aspects of the HES dataset which could limit its predictive ability. Hospital admissions prior to the introduction of computers in NHS hospitals were recorded on paper and later added to HES, which could imply greater risk of missingness or misclassification for older events. The use of hospital admissions data also misses less serious events which might still have predictive power for malignant melanoma.

Just over 94% of UKB participants are white, compared to around 86% in England and Wales [19]. This means that our model is most applicable to white populations, and may not generalise to the whole UK population. In particular, this model will not generalise well to ethnic minorities, who are still at risk of developing melanoma. The deprivation index measure that we included in our separate logistic regressions showed that an increase in wealth lead to an increased odds of melanoma, and, again it is unclear if this is generalise-able to the wider UK population, as the UKB population are on average wealthier and less deprived than the rest of the UK [17].

Currently, our models require a large amount of data to predict the occurrence of melanoma, making it difficult to implement quickly in a clinical setting. A possible clinical use of our model would be to assess immediate risk of developing melanoma, based on the previous 10-year medical history. Further extension for these modelling methods include reducing the time window for medical history and shifting the time window to be a certain time period before the melanoma occurs, to more of an early warning prediction. In our study, the window went up to the last hospital episode before melanoma, so an earlier prediction would be more clinically useful.

Other approaches to deep learning in prediction of skin malignancies have been focused on non-melanoma skin cancers [20]. Convolutional neural networks for modelling sequential data containing diagnostic information and medication prescriptions was found by Wang et al. [21] to be highly predictive (AUC 0.86-0.89) in predicting a 1-year incident risk of non-melanoma skin cancers. This approach could be employed for the prediction of the melanoma subtype in future studies. To address the 'black box' nature of neural networks,

further examination of the hidden state input-weights and output-weights with data pertubation as suggested by Olden et al. [22] could be done to identify variable importance. Since the model in use in this study is a shallow network with one hidden layer, singular neuronal ablation could be applied to observe its importance on predictive accuracy and inferring variable importance [23].

### Implications for Future Research

The time windows for occurrence of comorbidities in our studies went up to the day before melanoma diagnosis. To extend this model for early warning prediction, it would be recommended to shift windows for comorbidities further before the occurrence of melanoma, for example 1-3 years before a diagnosis. This would change the use of a predictive model to give early warning signals or to be preventative. Another avenue for further research would be to develop a risk score, by doing more extensive feature selection prior to fitting predictive models.

Our study found that using data from hospital admissions did not lead to huge increases in predictive accuracy. This suggests that the types of incidents that lead to hospitalisation were not strongly associated with the likelihood of developing melanoma. It is possible that events not requiring hospitalisation (such as fatigue, congenital nevi, mental health issues) could be better predictors of the development of malignant melanoma. An interesting avenue for future research would be to use records from primary care, eg. Clinical Practice Research Datalink (CPRD), to predict occurence of melanoma and associated risk.

## Author contributions statement

Authors 01206630, 01989284, 01986137, 02006075, and 01906532 had full access to the data and take responsibility for the integrity of the data and accuracy of data analysis. All authors contributed to study design formulation, data processing, analysis and interpretation and write-up of results Critical revision of manuscript, concept, study design and supervision: D. Vukovic, B. Metzler and M. Whitaker.

## Acknowledgments

The authors thank the the whole TDS and HDA teaching team for the support and advice given in the preparation of this report. We are appreciative of the reviewers comments that were received and addressed before submission of the final report. Responses were fully addressed in the supplementary material.

## References

1. Xuexia Wang, Michael J Oldani, Xingwang Zhao, Xiaohui Huang, and Dajun Qian. A Review of Cancer Risk Prediction Models with Genetic Variants. *Cancer Informatics*, 13(Suppl 2):19–28, September 2014. ISSN 1176-9351. doi: 10.4137/CIN.S13788. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4179686/`.
2. Maxwell Salvatore, Lauren J. Beesley, Lars G. Fritsche, David Hanauer, Xu Shi, Alison M. Mondul, Celeste Leigh Pearce, and Bhramar Mukherjee. Phenotype risk scores (PheRS) for pancreatic cancer using time-stamped electronic health record data: Discovery and validation in two large biobanks. *Journal of Biomedical Informatics*, 113:103652, January 2021. ISSN 1532-0464. doi: 10.1016/j.jbi.2020.103652. URL `https://www.sciencedirect.com/science/article/pii/S153204642030280X`.
3. Nasim Mavaddat, Kyriaki Michailidou, Joe Dennis, Michael Lush, Laura Fachal, Andrew Lee, Jonathan P. Tyrer, Ting-Huei Chen, Qin Wang, Manjeet K. Bolla, Xin Yang, Muriel A. Adank, Thomas Ahearn, Kristiina Aittomäki, Jamie Allen, Irene L. Andrulis, Hoda Anton-Culver, Natalia N. Antonenkova, Volker Arndt, Kristan J. Aronson, Paul L. Auer, Päivi Auvinen, Myrto Barrdahl, Laura E. Beane Freeman, Matthias W. Beckmann, Sabine Behrens, Javier Benitez, Marina Bermisheva, Leslie Bernstein, Carl Blomqvist, Bogdanova, and others. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *American Journal of Human Genetics*, 104(1): 21–34, January 2019. ISSN 1537-6605. doi: 10.1016/j.ajhg.2018.11.002.
4. Lars G. Fritsche, Lauren J. Beesley, Peter VandeHaar, Robert B. Peng, Maxwell Salvatore, Matthew Zawistowski, Sarah A. Gagliano Taliun, Sayantan Das, Jonathon LeFaive, Erin O. Kaleba, Thomas T. Klumpner, Stephanie E. Moser, Victoria M. Blanc, Chad M. Brummett, Sachin Kheterpal, Gonçalo R. Abecasis, Stephen B. Gruber, and Bhramar Mukherjee. Exploring various polygenic risk scores for skin cancer in the phenomes of the Michigan genomics initiative and the UK Biobank with a visual catalog: PRSWeb. *PLoS genetics*, 15(6): e1008202, June 2019. ISSN 1553-7404. doi: 10.1371/journal.pgen.1008202.
5. Caroline Elizabeth Mccarthy, Laura Jayne Bonnet, Michael Williams Marcus, and John K. Field. Development and validation of a multivariable risk prediction model for head and neck cancer using the UK Biobank. *International Journal of Oncology*, 57(5):1192–1202, November 2020. ISSN 1791-2423. doi: 10.3892/ijo.2020.5123.
6. Cancer Research UK. Melanoma skin cancer survival statistics, May 2015. URL `https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer/survival`.
7. Royal College of Pathologists. Meeting Pathology Demand - Histopathology Workforce Census. Technical report, Royal College of Pathologists, August 2018. URL `https://www.rcpath.org/uploads/assets/952a934d-2ec3-48c9-a8e6e00fcdca700f/Meeting-Pathology-Demand-Histopathology-Workforce-Census-2018.pdf`.
8. UK National Screening Committee. Screening in the UK: making effective recommendations 1 April 2018 to 31 March 2019. Technical report, August 2019. URL `https://www.gov.uk/government/publications/uk-national-screening-committee-recommendations-annual-report/screening-in-the-uk-making-effective-recommendations-1-april-2018-to-31-march-2019`.
9. David Whiteman. Predicting melanoma risk: theory, practice and future challenges. *Melanoma Management*, 1(2):105–114, November 2014. ISSN 2045-0885. doi: 10.2217/mmt.14.15. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6094627/`.
10. Catherine M. Olsen, Rachel E. Neale, Adèle C. Green, Penelope M. Webb, null The QSkin Study, null The Epigene Study, and David C. Whiteman. Independent validation of six melanoma risk prediction models. *The Journal of Investigative Dermatology*, 135(5):1377–1384, May 2015. ISSN 1523-1747. doi: 10.1038/jid.2014.533.

11. Isabelle Kaiser, Annette B. Pfahlberg, Wolfgang Uter, Markus V. Heppt, Marit B. Veierød, and Olaf Gefeller. Risk Prediction Models for Melanoma: A Systematic Review on the Heterogeneity in Model Development and Validation. *International Journal of Environmental Research and Public Health*, 17(21):7919, January 2020. doi: 10.3390/ijerph17217919. URL https://www.mdpi.com/1660-4601/17/21/7919. Number: 21 Publisher: Multidisciplinary Digital Publishing Institute.
12. Katarina Volkovova, Dagmar Bilanicova, Alena Bartonova, Silvia Letašiová, and Maria Dusinska. Associations between environmental factors and incidence of cutaneous melanoma. Review. *Environmental Health*, 11(Suppl 1): S12, June 2012. ISSN 1476-069X. doi: 10.1186/1476-069X-11-S1-S12. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3388446/.
13. Miriam Potrony, Celia Badenas, Paula Aguilera, Joan Anton Puig-Butille, Cristina Carrera, Josep Malvehy, and Susana Puig. Update in genetic susceptibility in melanoma. *Annals of Translational Medicine*, 3(15), September 2015. ISSN 2305-5839. doi: 10.3978/j.issn.2305-5839.2015.08.11. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4583600/.
14. Stephanie Carr, Christy Smith, and Jessica Wernberg. Epidemiology and risk factors of melanoma. *Surgical Clinics of North America*, 100(1):1–12, 2020. ISSN 0039-6109. doi: https://doi.org/10.1016/j.suc.2019.09.005. URL https://www.sciencedirect.com/science/article/pii/S0039610919301239. Contemporary Melanoma Management: A Surgical Perspective.
15. R. De Maesschalck, D. Jouan-Rimbaud, and D.L. Massart. The mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50(1):1–18, 2000. ISSN 0169-7439. doi: https://doi.org/10.1016/S0169-7439(99)00047-7. URL https://www.sciencedirect.com/science/article/pii/S0169743999000477.
16. Jerome Pages. *Multiple Factor Analysis by Example Using R*. Chapman and Hall/CRC, Canada, November 2014. ISBN 978-1-4822-0547-3.
17. Anna Fry, Thomas J Littlejohns, Cathie Sudlow, Nicola Doherty, Ligia Adamska, Tim Sprosen, Rory Collins, and Naomi E Allen. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *American Journal of Epidemiology*, 186(9):1026–1034, 06 2017. ISSN 0002-9262. doi: 10.1093/aje/kwx246. URL https://doi.org/10.1093/aje/kwx246.
18. Sofie A.E. De Hertog, Christianne A.H. Wensveen, Maarten T. Bastiaens, Christine J. Kielich, Marjo J.P. Berkhout, Rudi G.J. Westendorp, Bert J. Vermeer, and Jan N. Bouwes Bavinck and. Relation between smoking and skin cancer. *Journal of Clinical Oncology*, 19(1):231–238, January 2001. doi: 10.1200/jco.2001.19.1.231. URL https://doi.org/10.1200/jco.2001.19.1.231.
19. GOV.UK. Population of england and wales. URL https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/population-of-england-and-wales/latest.
20. David Roffman, Gregory Hart, Michael Girardi, Christine J. Ko, and Jun Deng. Predicting non-melanoma skin cancer via a multi-parameterized artificial neural network. *Scientific Reports*, 8(1):1701, January 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-19907-9. URL https://www.nature.com/articles/s41598-018-19907-9. Number: 1 Publisher: Nature Publishing Group.
21. Hsiao-Han Wang, Yu-Hsiang Wang, Chia-Wei Liang, and Yu-Chuan Li. Assessment of Deep Learning Using Nonimaging Information and Sequential Medical Records to Develop a Prediction Model for Nonmelanoma Skin Cancer. *JAMA Dermatology*, 155(11):1277, November 2019. ISSN 2168-6068. doi: 10.1001/jamadermatol.2019.2335. URL https://jamanetwork.com/journals/jamadermatology/fullarticle/2749356.
22. Julian D Olden, Michael K Joy, and Russell G Death. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*, 178(3-4):389–397, November 2004. ISSN 03043800. doi: 10.1016/j.ecolmodel.2004.03.013. URL https://linkinghub.elsevier.com/retrieve/pii/S0304380004001565.
23. Richard Meyes, Melanie Lu, Constantin Waubert de Puiseau, and Tobias Meisen. Ablation Studies in Artificial Neural Networks. *arXiv:1901.08644 [cs, q-bio]*, February 2019. URL http://arxiv.org/abs/1901.08644. arXiv: 1901.08644.

## Appendix I: Prevalence of Comorbidities

**Table 4.** Prevalent comorbidities derived from the UKB dataset, sorted according to descending frequencies.

| Comorbidities | Controls N (%) | Cases N (%) | All Participants N (%) |
|---|---|---|---|
| Hypertension | 3147 (29.83 %) | 1632 (30.94 %) | 4779 (30.2 %) |
| Hypercholestrolaemia | 1569 (14.87 %) | 747 (14.16 %) | 2316 (14.64 %) |
| Malignant Melanoma | 40 (0.38 %) | 1905 (36.11 %) | 1945 (12.29 %) |
| Asthma | 1228 (11.64 %) | 512 (9.71 %) | 1740 (11 %) |
| Osteoarthritis | 1098 (10.41 %) | 516 (9.78 %) | 1614 (10.2 %) |
| Hayfever/Allergic Rhinitis | 602 (5.71 %) | 287 (5.44 %) | 889 (5.62 %) |
| Depression | 619 (5.87 %) | 246 (4.66 %) | 865 (5.47 %) |
| Hypothyroidism/Myxoedema | 561 (5.32 %) | 247 (4.68 %) | 808 (5.11 %) |
| Unclassifiable | 539 (5.11 %) | 256 (4.85 %) | 795 (5.02 %) |
| Gastro-Oesophageal Reflux | 509 (4.82 %) | 251 (4.76 %) | 760 (4.8 %) |
| Diabetes | 551 (5.22 %) | 208 (3.94 %) | 759 (4.8 %) |
| Angina | 443 (4.2 %) | 178 (3.37 %) | 621 (3.92 %) |
| Migraine | 305 (2.89 %) | 174 (3.3 %) | 479 (3.03 %) |
| Hiatus Hernia | 279 (2.64 %) | 136 (2.58 %) | 415 (2.62 %) |
| Myocardial Infarction | 303 (2.87 %) | 103 (1.95 %) | 406 (2.57 %) |
| Breast Cancer | 241 (2.28 %) | 153 (2.9 %) | 394 (2.49 %) |
| Eczema/Dermatitis | 275 (2.61 %) | 109 (2.07 %) | 384 (2.43 %) |
| Irritable Bowel Syndrome | 267 (2.53 %) | 110 (2.09 %) | 377 (2.38 %) |
| Cataract | 248 (2.35 %) | 117 (2.22 %) | 365 (2.31 %) |
| Enlarged prostate | 224 (2.12 %) | 139 (2.64 %) | 363 (2.29 %) |
| Deep Venous Thrombosis (dvt) | 232 (2.2 %) | 117 (2.22 %) | 349 (2.21 %) |
| Uterine Fibroids | 189 (1.79 %) | 136 (2.58 %) | 325 (2.05 %) |
| Osteoporosis | 208 (1.97 %) | 93 (1.76 %) | 301 (1.9 %) |
| Back problem | 203 (1.92 %) | 98 (1.86 %) | 301 (1.9 %) |
| Pneumonia | 187 (1.77 %) | 105 (1.99 %) | 292 (1.85 %) |
| Prolapsed Disc/Slipped Disc | 187 (1.77 %) | 102 (1.93 %) | 289 (1.83 %) |
| Basal Cell Carcinoma | 104 (0.99 %) | 183 (3.47 %) | 287 (1.81 %) |
| Cholelithiasis/Gall stones | 198 (1.88 %) | 83 (1.57 %) | 281 (1.78 %) |
| Gout | 160 (1.52 %) | 93 (1.76 %) | 253 (1.6 %) |
| Stroke | 175 (1.66 %) | 73 (1.38 %) | 248 (1.57 %) |
| Anxiety/Panic attacks | 155 (1.47 %) | 70 (1.33 %) | 225 (1.42 %) |
| Glaucoma | 142 (1.35 %) | 75 (1.42 %) | 217 (1.37 %) |
| Emphysema/Chronic Bronchitis | 160 (1.52 %) | 56 (1.06 %) | 216 (1.36 %) |
| Diverticular disease/Diverticulitis | 143 (1.36 %) | 67 (1.27 %) | 210 (1.33 %) |
| Rheumatoid Arthritis | 137 (1.3 %) | 67 (1.27 %) | 204 (1.29 %) |
| Atrial Fibrillation | 137 (1.3 %) | 65 (1.23 %) | 202 (1.28 %) |
| Prostate Cancer | 127 (1.2 %) | 65 (1.23 %) | 192 (1.21 %) |
| Arthritis (nos) | 141 (1.34 %) | 50 (0.95 %) | 191 (1.21 %) |
| Chickenpox | 108 (1.02 %) | 72 (1.36 %) | 180 (1.14 %) |
| Psoriasis | 122 (1.16 %) | 53 (1 %) | 175 (1.11 %) |
| Tonsiltis | 106 (1 %) | 65 (1.23 %) | 171 (1.08 %) |
| Dermatological Cancer | 28 (0.27 %) | 141 (2.67 %) | 169 (1.07 %) |
| Allergy or Drug Anaphylaxis | 110 (1.04 %) | 54 (1.02 %) | 164 (1.04 %) |
| Sciatica | 110 (1.04 %) | 54 (1.02 %) | 164 (1.04 %) |
| Measles / Morbillivirus | 102 (0.97 %) | 59 (1.12 %) | 161 (1.02 %) |
| Appendicitis | 108 (1.02 %) | 51 (0.97 %) | 159 (1 %) |

## Appendix II: Neural Network Supplemental Figures

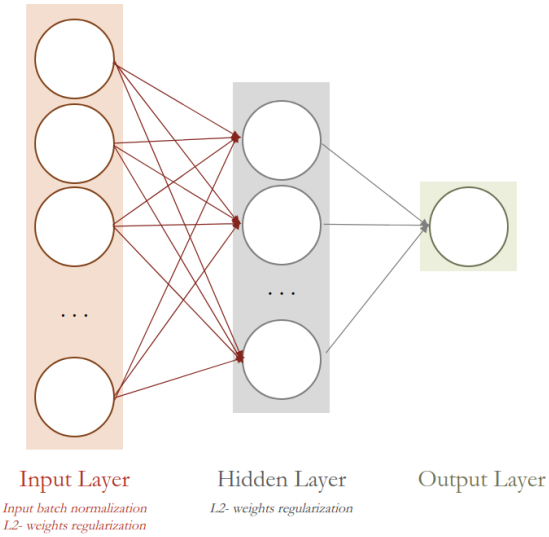
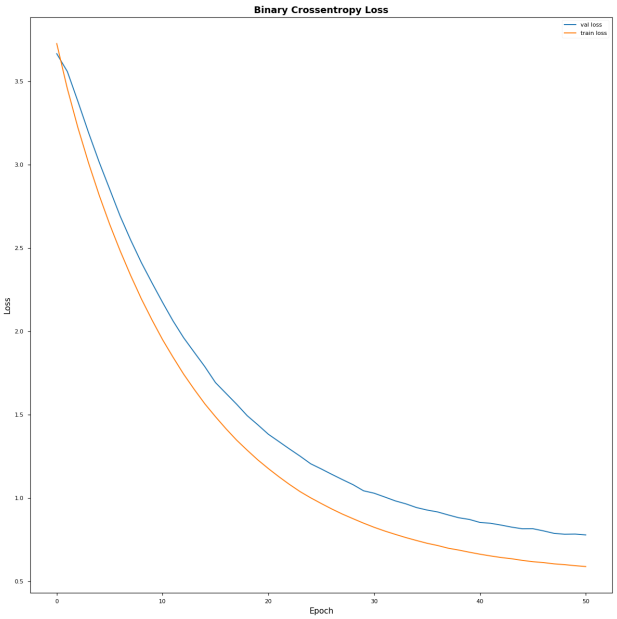
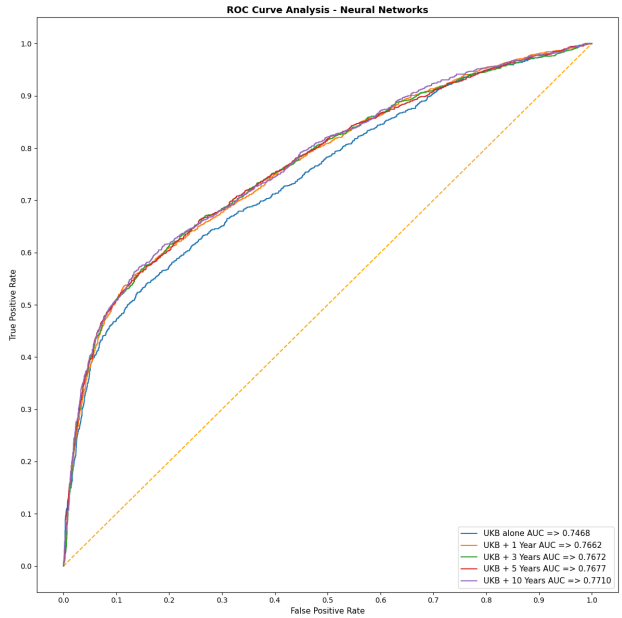



**Fig. 5.** Graphical depiction of best performing neural network architecture. Input layer was formed of 1570 neurons, with an input vector containing demographic information, biomarker measurements, and time-windowed comorbidities. Weights regularization was implemented on both the input and hidden layer using the L2-norm penalty with lambda 0.05. Hidden layer consisted of 32 fully connected neurons. The output layer was a single neuron that outputs prediction probability after sigmoid activation function was applied.



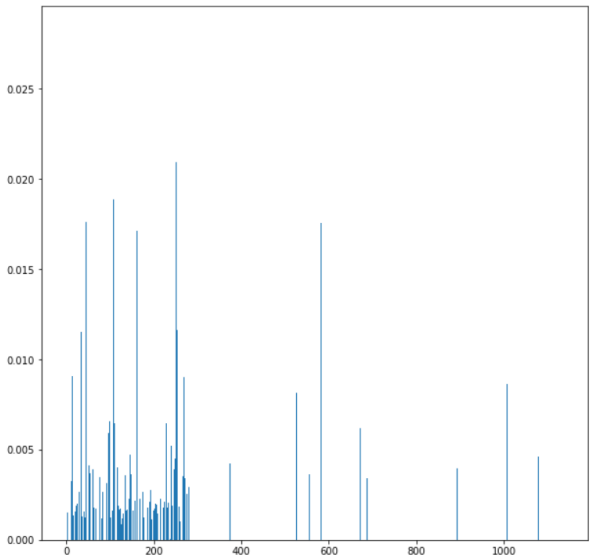

(a) Training and validation set binary crossentropy loss over each epoch. Learning rate = 0.001 with stochastic gradient descent.



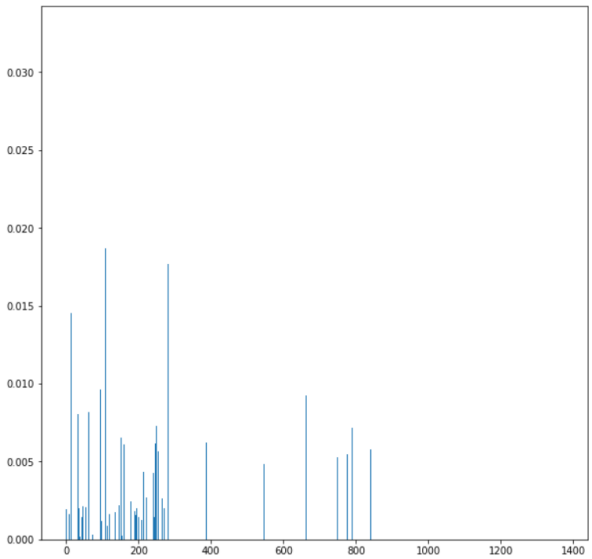

(b) Receiver Operator Curve graph of the best performing neural network model applied to various time-windowed comorbidities.

**Fig. 6.** Supplemental figures depicting training, validation and test dataset information. (a) shows the decrease of loss function on both training and validation datasets over time and model convergence at 51 epochs. To prevent overfitting, early stopping terminated further training of the neural network. (b) The blue line denotes AUC curve of using features from UKB alone, and the effects of addition of time windows on AUC.

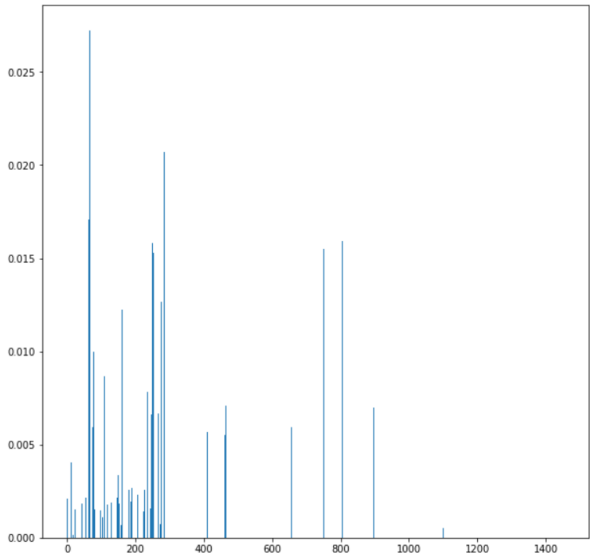
## Appendix III: XGBoost feature importance plots



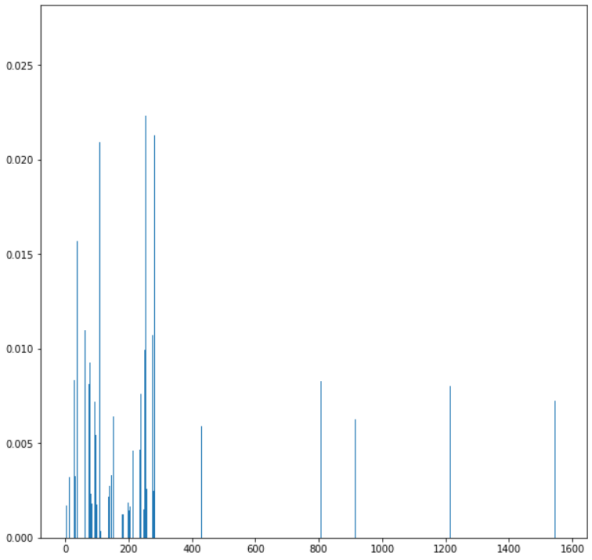

(a) Feature importance plots for 1 year time windowed plus UKB feature set XGBoost model.




(b) Feature importance plots for 3 year time windowed plus UKB feature set XGBoost model.




(c) Feature importance plots for 5 year time windowed plus UKB feature set XGBoost model.




(d) Feature importance plots for 10 year time windowed plus UKB feature set XGBoost model.

**Fig. 7.** Supplemental feature importance plots for the 4 different time windows used in XGBoost modelling.

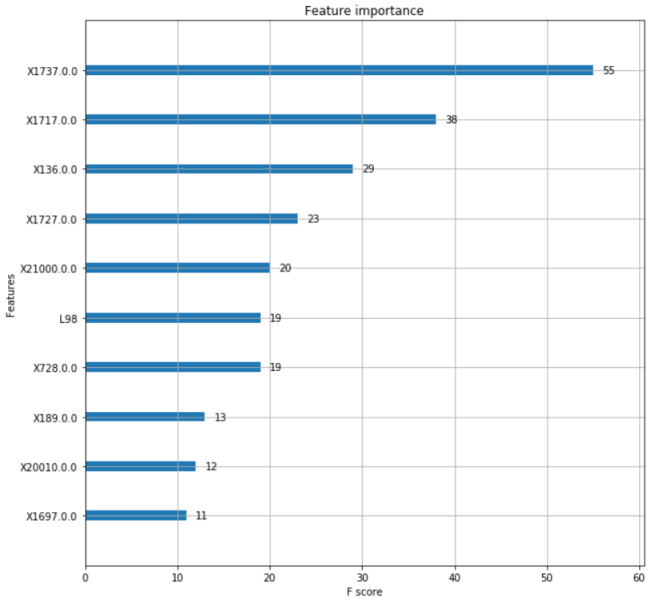



**Fig. 8.** The top ten most important features for prediction in the 1 year time windowed XGBoost model. Features beginning in "X" are from UKB feature set. Features with ICD codings are from time windowed phenotypical data. The top five features in order of importance are childhood sunburn occasions, skin colour, number of operations, ease of skin tanning and ethnicity.

## Response to Peer-review Comments

Our report was peer reviewed by prior to completion. Overall, greater elaboration on the process of obtaining the analysis datasets, including the process of time windowing, missing data handling, and methods of dimensionality reduction was performed as recommended and detailed below.

| Review Criteria | Reviewers' Comment | Response to Comments |
|---|---|---|
| Scientific justification of the work proposed | Opportunities identified from previous work: Are you planning to address a gap in the literature, or any limitations from previous studies, or is this more some replication work to a new cancer type? | Previous studies using phenotype risk scores were done on other cancer types as specified (head and neck; pancreatic cancers). Yes, this work aims to replicate this approach on malignant melanoma. |
| | Highlight the aims of this study | This was now reiterated in both the abstract and under the motivation section. |
| | Possibility of applying the model for predicting melanoma risk? | A section titled 'Implications for Future Research' was added. |
| | Lack of depth as to why machine learning methods can be used for cancer risk prediction. | Several additional reference articles were added to address this, refer to references number [2, 3, 4, 11]. This was also explained further under 'Introduction' and 'Motivation' sections. |
| Detailed description of analytical plan and methods used | "We identified groups of variables we deemed to have clinical relevance as potential covariates in our analyses." How were these identified? | Covariates were identified based on a priori clinical knowledge, literature search of environmental risk factors and confirmed via separate logistic regression analyses. Additional references were added [12, 13, 14]. |
| | Could you provide more details on the comorbidities on which your analysis is based on? Why some are regrouped by chapters, and others to more specific disease categories? | This was referring to the 'Follow-up Period for Hospital Admissions' subsection. Comorbidities were windowed according to ICD-10 standards, and detailed example of this process was given. |
| | Possible extensions: add an additional time window and compare results. | A total of five different time-windows were applied in the final analyses, instead of the initial three windows. This were windowed according to no comorbidities, 1-year, 3-years, 5-years and 10-years of comorbidities. |
| | Use numbers to specify size of cohort and of dataset. What are the specific 3 FAMD components that were selected? | The numbers of cases and controls were specified. We attempted both FAMD and PCA for dimensionality reduction techniques on the dataset. Due to technical issues with the FAMD package in Python we and decided on PCA for further analyses. The number of components selected and proportion of explained variance was detailed under the subsection 'Dimensionality Reduction'. |
| | "We ran univariate logistic regression models the each of these key confounding factors, adjusted for age and gender only." Assumptions for logistic regression met? | Yes, logistic regression assumptions were met. Each observation was an independent measure of a unique individual. |
| | Perhaps some more information about mahalanobis distance metric? | This is a well-defined metric, additional description was given under the subsection 'Study Population' and further information can be obtained from reference number [15]. |
| | What other data cleaning or data manipulation steps were performed prior to the imputation? | Features with potential for data leakage due to the time stamped approach of analysis were removed. Also, features that had over 40% missingness were also excluded. Full details of the data cleaning and data manipulation is described in the section 'Data Preparation'. |

( To be continued)

| Review Criteria | Reviewers' Comment | Response to Comments |
|---|---|---|
| | "As the aim of this study was simply to understand what level of prediction can be achieved from comorbidity data – as opposed to aiming to create a diagnosis rubric for clinicians – we believed that an ensemble method was valid and appropriate". Isn't this in contradiction with what specified in the motivation section? | The motivation and aims were reiterated in both the abstract and 'Motivation & Aims' section to provide additional clarity and solidify the objective of the study. Yes, the initial motivation section was unclear and this was addressed in the second comment from the reviewers. |
| In-depth presentation and interpretation of results | Random Forest - assess feature importance | Please see appendix III and XGBoost results section for XGBoost feature importance plots. These are broadly the same features that were selected by random forest. |
| | Model comparison - predictive model comparison to any sort of baseline model with clinically relevant covariates only? | No, this study aimed to explore how environmental risk factors in relation to time-windowed comorbidties could provide predictive accuracy to machine learning models. |
| | Do you plan to "reverse engineer" results to analyse/compare/interpret relevant comorbidities between the different models? | No, machine learning techniques were applied to identify comorbidities that were most predictive of malignant melanoma, for example from the variable importance plots of tree-based models. This study approaches the comorbidities from a agnostic standpoint. |
| | Neural Network - 10000 epochs are an unusually high number. How do you justify this decision? | Yes, this was a large number of epochs that lead to overfitting on the dataset. Subsequent analysis performed utilized early stopping callbacks and limited the upper number of epochs to a maximum of 300 to 2000 epochs. |
| | Do you plan to include measures of training and test set loss for tree-based methods? | The training and test set loss was manually monitored for each iteration of the tree-based methods. This was done by adjusting the optimal hyperparameters identified after a grid search and noting the difference in the training and test set loss and accuracy metrics. For example, increasing the depth parameter beyond four lead to a larger difference between the training and test set accuracy's - signalling overfitting. Due to the large hyperparameter space explored, we reported the best performing models only. |
| Identification of the added value of results and critical appraisal of work | Limited clinical applications? There is an added value in predicting the risk of cancer years before it occurs, less so in using 'immediate risk'. Good proof of concept for using comorbidities in melanoma risk prediction, but other timeframes with more clinically relevant applications could be considered. This could help align the work with the scientific justification provided, particularly regarding the need for an early screening. You could consider shifting your study aims of providing a 'clinically useful predictive model for individual level melanoma risk' to something more aligned with the methods chosen. | We agreed with the reviewers' comment and the final report aligns the aims with the use of time-windows and the methods chosen. Further comments on the potential use of a 10-year window for accessing immediate risk of developing melanoma was mentioned in under 'Discussion' and a 'Future Research' section was added to address this point. |
| | Why do you emphasize the low false negative rate? In the specific context, isn't it more important to focus on maintaining a low false positive rate? | Missing a malignant melanoma diagnosis was much more detrimental due to the ethical reasons, potential harm to the patient in the event of disease progression and not to mention medicolegal implications. The risks of false positive reporting of melanoma on the other hand is simply the increase in clinical burden but does not carry the implications mentioned above. |
| | Potential drawbacks of study, limitations of models used, critiques about dataset are not mentioned | These sections were added in the final report, under the 'Discussion' section. |