




Article

Integration of lncRNAs, Protein-Coding Genes and Pathology Images for Detecting Metastatic Melanoma

Shuai Liu ¹ , Yusi Fan ², Kewei Li ¹, Haotian Zhang ¹ , Xi Wang ¹, Ruofei Ju ¹, Lan Huang ¹, Meiyu Duan ¹ and Fengfeng Zhou ^{1,*} 

¹ College of Computer Science and Technology, and Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China

² College of Software, and Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China

* Correspondence: ffzhou@jlu.edu.cn; Tel./Fax: +86-431-8516-6024

Abstract: Melanoma is a lethal skin disease that develops from moles. This study aimed to integrate multimodal data to predict metastatic melanoma, which is highly aggressive and difficult to treat. The proposed EnsembleSKCM method evaluated the prediction performances of long noncoding RNAs (lncRNAs), protein-coding messenger genes (mRNAs) and pathology images (images) for metastatic melanoma. Feature selection was used to screen for metastatic biomarkers in the lncRNA and mRNA datasets. The integrated EnsembleSKCM model was built based on the weighted results of the lncRNA-, mRNA- and image-based models. EnsembleSKCM achieved 0.9444 in the prediction accuracy of metastatic melanoma and outperformed the single-modal prediction models based on the lncRNA, mRNA and image data. The experimental data suggest the importance of integrating the complementary information from the three data modalities. WGCNA was used to analyze the relationship of molecular-level features and image features, and the results show connections between them. Another cohort was used to validate our prediction.

Keywords: bioinformatics; metastatic melanoma; multimodal information fusion; lncRNA; feature selection



Citation: Liu, S.; Fan, Y.; Li, K.; Zhang, H.; Wang, X.; Ju, R.; Huang, L.; Duan, M.; Zhou, F. Integration of lncRNAs, Protein-Coding Genes and Pathology Images for Detecting Metastatic Melanoma. *Genes* **2022**, *13*, 1916. <https://doi.org/10.3390/genes13101916>

Academic Editors: Jinchen Li, Chao Chen and Mingbang Wang

Received: 5 September 2022

Accepted: 18 October 2022

Published: 21 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Melanoma is a type of malignant skin cancer, and its incidence rate has increased rapidly in recent decades [1–3]. The melanoma incidence rate in Canada is 122.9 cases per million person a year [4]. The melanoma mortality rate in the U.S. increased by 7.5% from 1986 to 2013 [5]. Early detection and appropriate treatment of non-metastatic melanoma may decrease the mortality rate and substantially increase survival [6].

Computational models are widely used in the field of disease diagnosis and prognosis. Systems biology models can incorporate various data sources such as mechanistic details of biological mechanisms, inter-patient variability and drug–target interactions into the translational research [7]. Verma et al. fine-tuned the model to predict the liver regeneration process by integrating signaling mechanisms and cellular functional state transitions [8]. Compared with other studies, they performed the liver failure classification to characterize the response of recovery and failure. Milberg et al. developed a quantitative systems pharmacology (QSP) model for the combination immunotherapy specific to melanoma [9]. With the development of high-throughput sequencing and biological technologies, a huge amount of biomedical data is being rapidly accumulated and machine learning approaches are also actively utilized.

The detection of metastatic melanoma using machine learning algorithms is a clinically useful and computationally challenging task. Several machine learning methods delivered promising prediction performances for metastatic melanoma. Bellomo et al. employed classifier logistic regression to combine clinicopathologic and gene expression data for

the detection of metastatic melanoma in sentinel lymph nodes [10]. Garg et al. trained a random forest model with the screened signature genes to detect metastatic melanoma [11]. Mancuso et al. predicted metastasis in melanoma patients with high and low risk of metastasis by serum cytokines and Breslow thickness [12]. Shepelin et al. used the SVM algorithm to identify 44 characteristic signaling pathways associated with melanoma metastasis [13].

Melanoma may be detected by machine learning methods. Melanoma develops from moles, and its early detection through the ABCDE criteria may generate some false negatives depending on the experience of the practicing dermatologists [14]. There are two main categories of machine learning-based melanoma detection methods, i.e., image-based and OMIC-based methods [15–18]. Adytia et al. proposed a novel transfer learning method to classify skin lesions based on the internet of health things [19]. In addition to the classification task, lesion segmentation is another important machine learning task for detecting melanoma [20–22]. Rasmiranjan et al. optimized a set of hyperparameters of a fully convolutional encoder–decoder network (FCEDN) to segment skin cancer lesions [23]. Tang et al. employed end-to-end multistage UNets to segment skin lesions accurately [24]. Afsah et al. explored the feasibility of using hybrid textural analysis to segment and classify skin cancers based on dermoscopic images [25].

OMIC data provide a molecular-level view of melanoma, and the detected biomarkers facilitate the understanding of the onset and progression mechanisms of melanoma. Lai et al. selected the fully connected melanoma subnetwork with the best modularity score and proposed an autoencoder-based deep learning network to detect different melanoma subgroups using the genomic data in The Cancer Genome Atlas [26]. Wei revealed 798 differentially expressed genes of melanoma and built a support vector machine (SVM)-based classifier using the top 110 biomarker genes to achieve at least 0.944 in accuracy across three independent datasets [27]. Aigli et al. proposed an ensemble dimensionality reduction technique to estimate melanoma patient prognosis in a large cohort [28].

Long noncoding RNAs (lncRNAs) may act as competing endogenous RNAs (ceRNAs) and are widely involved in tumor onset and progression [29]. Multiple previous investigations have demonstrated that lncRNAs have a close relationship with the prognosis of melanoma [30–32]. Yan et al. screened 61 lncRNAs associated with melanoma prognosis and built a weighted risk score model based on seven key candidate lncRNAs [33]. The lncRNA U731166 was also observed to be upregulated during the migration and invasion of melanoma and even played a role in developing vemurafenib resistance [34].

Both image and OMIC data provide complementary information about melanoma, and the integrated analysis of multimodal data is important to fully utilize these different data modalities [35,36]. Both image and OMIC data have large numbers of features, and most do not contribute to detecting melanoma. Such “large p small n” datasets may lead to prediction model overfitting [37]. Feature selection is one of the methods used to tackle this challenge [38–40].

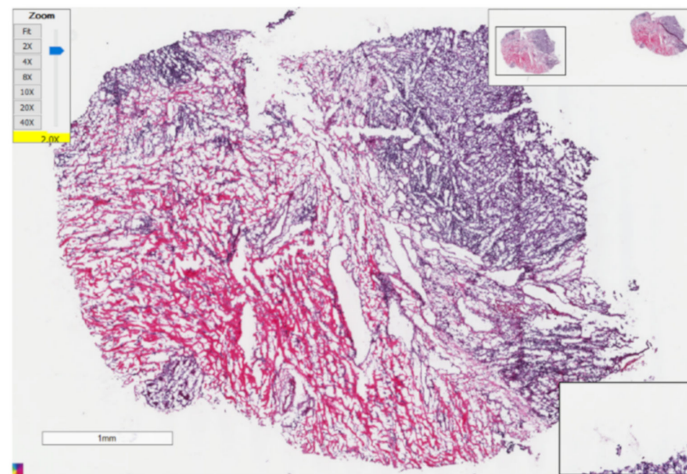
This study proposes an ensemble detection algorithm, EnsembleSKCM, for metastatic melanoma by integrating the data sources of lncRNAs, protein-coding mRNAs and pathology images. The extremely high dimensionalities of lncRNAs and mRNAs were screened for redundancies by feature selection algorithms. The image-extracted cell features were combined with the results of lncRNAs and mRNAs for the final classification between metastatic and non-metastatic melanoma samples. The experimental data support the necessity of integrating multimodal data for the detection of metastatic melanoma. The Python source code and the multi-modal datasets are freely available at <http://www.healthinformatics.org/supp/>, accessed on 4 September 2022.

2. Materials and Methods

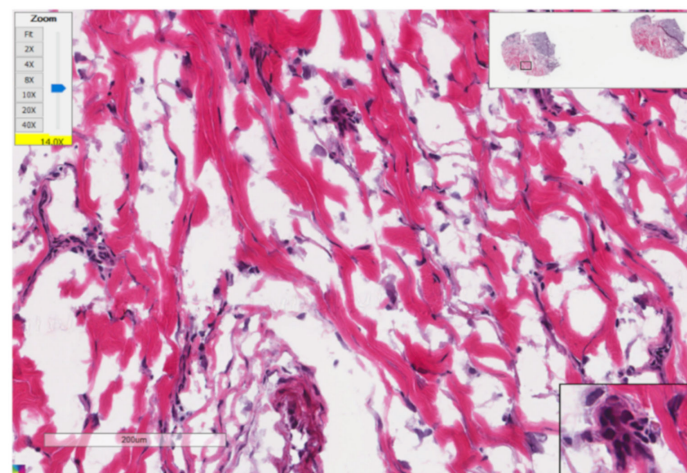
2.1. Summary of Datasets

This study retrieved the features of pathology images, lncRNAs and mRNAs of melanoma from The Cancer Genome Atlas (TCGA) database [41,42]. The mRNA expression refers to the expression level of the corresponding gene. FPKM was used to normalize the transcript expression data. The dataset's metadata are shown in Supplementary Table S1. This TCGA-SKCM cohort selected patients with a diagnosis of primary metastatic cutaneous melanoma or metastatic melanoma of an unknown primary, and they were also required to have had no previous systemic therapy (except that adjuvant interferon- $\alpha \geq 90$ days prior was permitted) [43]. This study investigated the integrated analysis of metastatic melanoma based on multi-modality data sources, i.e., lncRNA, mRNA and pathology images. Therefore, the samples without regional lymph node metastasis information were removed. There were 414 melanoma samples with regional lymph node metastasis information, lncRNAs, mRNAs and images in the TCGA database. The number of primary melanoma samples was 235 and that of metastasis melanoma samples was 179. Multiple samples may be extracted from one patient and one of these samples from the same patient was randomly chosen for further analysis. There were 411 remaining patients, among whom there were 257 males and 154 females. The majority (392) of this cohort was white people, and there were only 12 Asian and 1 black or African American. This cohort consisted of only 19 patients under the age of 30. The remaining patients included 191 and 183 patients under and over 60-years-old, respectively. Some samples did not have information on sex, ethics or age.

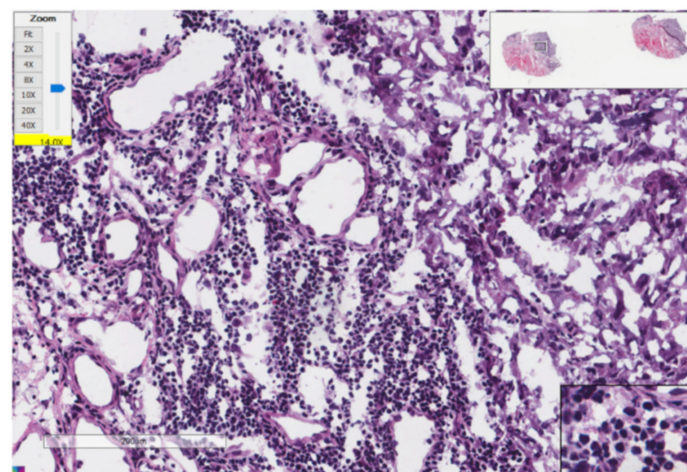
A binary classification between primary melanoma ($n = 0$, positive samples) and melanoma with regional lymph node metastasis ($n > 0$, negative samples) was investigated. A sample was a melanoma patient, and a feature was an lncRNA's expression level, an mRNA's expression level or a cell type's percentage within the pathological image in this study. Each sample had 6919 lncRNA features and 19051 mRNA features. A feature in a sample was the expression level of the corresponding lncRNA or mRNA gene, or the percentage of a cell type in the pathological image of that sample. The annotations were generated by UCSC for the Dec. 2013 (GRCh38/hg38) assembly of the human genome. For the processing of missing values, all of the 411 samples were considered (including both primary and lymph node metastatic melanoma samples). If the ratio of missing values of a feature was greater than 50%, this feature was removed. The remaining missing values were filled with 0, assuming that the sequencing technology cannot detect the low expression levels of these genes. We did not process the missing values in the primary and metastatic melanoma samples separately, to avoid the case that these two groups of samples may have different features. Eight features were retrieved from the pathology images to describe the percentages of lymphocyte infiltration, monocyte infiltration, necrosis, neutrophil infiltration, normal cells, stromal cells, tumor cells and tumor nuclei. The detection and counting of the different cell types were conducted using the CellProfiler software and the percentages of these cell types were calculated as the representative features for the pathology images [44]. After the preprocessing step, there were 1716 lncRNAs, 1827 mRNAs and 8 image features. The example pathology images are shown in Figure 1 by the freeware ImageScope version 12.4.6 [45]. The image came from the sample "TCGA-BF-A1Q0-01A-02-TSB". The scale bar of Figure 1a is 1 mm and the scale bars of Figure 1b,c are 200 μm . Figure 1b,c were cropped randomly from (a). This dataset was denoted as "TCGA-SKCM".



(a)



(b)



(c)

Figure 1. The example pathology image and its randomly cropped patches. (a) Shows the whole pathology image; (b,c) are patch images cropped randomly from (a) to show the detailed information.

We screened the GEO database and found only one transcriptomic cohort (GSE59455) of metastatic melanoma for further validation of the experimental results in the above sections. We did not find a cohort with both pathological images and RNA-seq transcrip-

tomes. The GSE59455 dataset was an array-based transcriptomic profile and consisted of 141 samples. It does not have as detailed metadata as the TCGA–SKCM dataset. The samples without information on primary or metastatic cancers were removed, and the remaining samples consisted of 17 primary cancers and 43 metastatic cancers. The array-based GSE59455 dataset and the RNA-seq-based transcriptomic dataset from the TCGA database had large differences in both the feature list and the expression patterns. Only 2 lncRNAs and 198 mRNAs overlapped between the GSE59455 and the TCGA melanoma datasets, while the optimal EnsembleSKCM model used 200 lncRNAs and 200 mRNAs from the RNA-seq-based transcriptomic profiles.

2.2. Performance Measurements

The classification model was evaluated by five widely used measurements, i.e., accuracy, F1-score, precision, recall and AUC. Assume that P and N are the numbers of positive and negative samples, respectively. The numbers of correctly predicted positive and negative samples are true positive (TP) and true negative (TN). False positive (FP) and false negative (FN) are the numbers of incorrectly predicted positive and negative samples. The performance measurements were defined as follows.

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

AUC is the area under the ROC curve, which is a good parameter-independent measurement for a binary classification model. A stratified 5-fold cross-validation (S5FCV) strategy was used to evaluate the models.

2.3. The Proposed EnsembleSKCM Method

The proposed EnsembleSKCM algorithm integrated the information of lncRNAs, mRNAs and pathology images to detect metastatic melanoma, as shown in Figure 2. A data preprocessing step was used to remove the features, and an additional step of feature selection was used to remove the redundant lncRNA and mRNA features to avoid the overfitting problem [46–48]. For image data, a feature extraction step was used to obtain the image feature representation. A three-layer fully connected neural network was designed to detect metastatic melanoma using eight pathology image-based features. The three data sources were then ensembled with different weights to generate the final prediction results.

2.4. Feature Selection and Classification Algorithms

Three feature selection algorithms were evaluated in this study, including SVM–RFE [49], variance [50] and t-test [51]. SVM–RFE trained a support vector machine (SVM) model [52,53] and selected features using their weights in the trained SVM model to alleviate the possibility of the “large p small n” paradigm [37]. The redundant features were iteratively removed by the SVM-based recursive feature elimination (SVM–RFE) strategy [49,54,55]. The incremental feature selection (IFS) strategy [56] was used to find the best subset of features ranked by variance (descendent order) or t-test (ascendent order).

This study evaluated six classifiers in the prediction task of metastatic melanoma: random forest (RF) [57], support vector machine (SVM) [58], linear regression (LR) [59], k-nearest neighbor (KNN) [60], decision trees (DT) [61] and naïve Bayes (NB) [62].

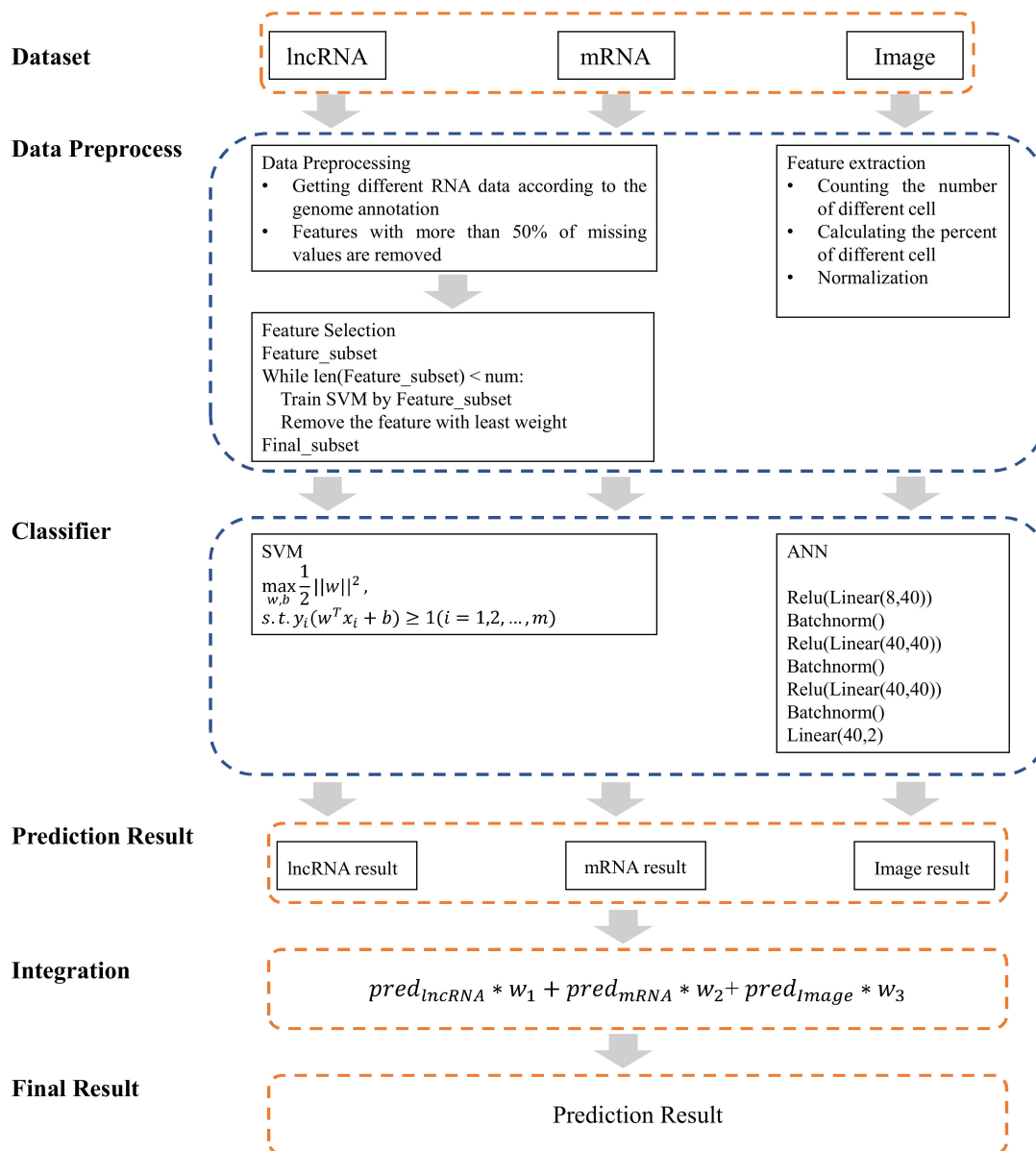


Figure 2. The workflow of the proposed EnsembleSKCM. First, for lncRNAs and mRNAs, data preprocessing was used to remove missing values, and feature selection methods were used to remove redundant features. For images, feature extraction was used to obtain structural features. Next, a support vector machine (SVM) was used to predict metastatic melanoma for lncRNAs and mRNAs, and an artificial neural network (ANN) was used to predict metastatic melanoma for imaging features. Finally, the results from different data modalities were integrated by assigning different weights.

2.5. Fully Connected Neural Network

A two-layer fully connected neural network was designed to predict metastatic melanoma using pathology image features, as shown in Figure 2. The first layer was designed as $X_1 = Batchnorm(relu(linear(X_0)))$, while the second layer was designed as $X_2 = Batchnorm(relu(linear(X_1)))$. The third layer was designed as $X_3 = Batchnorm(relu(linear(X_2)))$. The output layer was designed as $X_4 = Batchnorm(relu(linear(X_3)))$. X_0 is the input data of the network, and X_4 is the output data of the network.

2.6. Construction of the WGCNA Network

To verify whether image features contribute useful information to the molecular features, the WGCNA analysis was used to calculate the relationship between these two kinds of features. If the correlation between two features was high, these features were redundant to each other. Otherwise, the image features represented useful information for predicting metastatic melanoma. A total of 400 lncRNAs and mRNAs were screened out by the WGCNA package version 1.70–3 [63] in the R-Studio 4.1.3 software.

2.7. Implementation Details

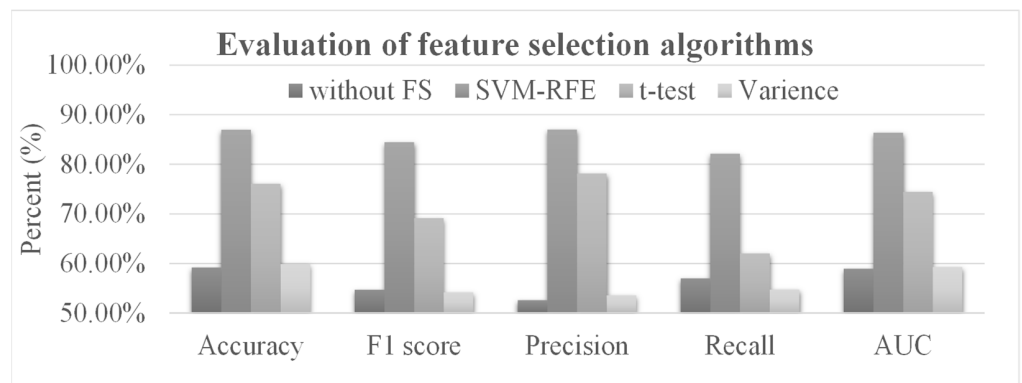
The proposed EnsembleSKCM framework integrated multimodal information to predict metastatic melanoma. For lncRNA features, SVM-RFE was used to remove the inter-feature redundancy. The SVM used the linear kernel, and the parameter C was 0.1. SVM-RFE removed one feature per iteration. For mRNA features, SVM-RFE was used to remove redundant features again, and the linear kernel was used. The parameter C was 1. SVM-RFE removed 10 features per iteration due to the very large number of mRNA features. Then, the SVM classifier was used to classify the metastatic version of non-metastatic melanoma. The parameter C was evaluated by the values 0.1, 1, 10 and 100. The four kernels were evaluated, including 'linear', 'rbf', 'poly' and 'sigmoid'. The grid search was used to screen for the best parameter values. For the image features, we built a two-hidden-layer neural network. Both hidden layers contained 40 neurons. The stochastic gradient descent (SGD) with a batch size of 32 was used to optimize our model for 1000 epochs. The momentum and weight decay parameters were set to 0.9 and 1×10^{-4} , respectively. The initial learning rate was 0.01. This study was carried out on the Windows 10 operating system with an Intel(R) Core(TM) i7–8750H CPU@2.20GHZ 2.21GHZ and 8 GB RAM.

3. Results

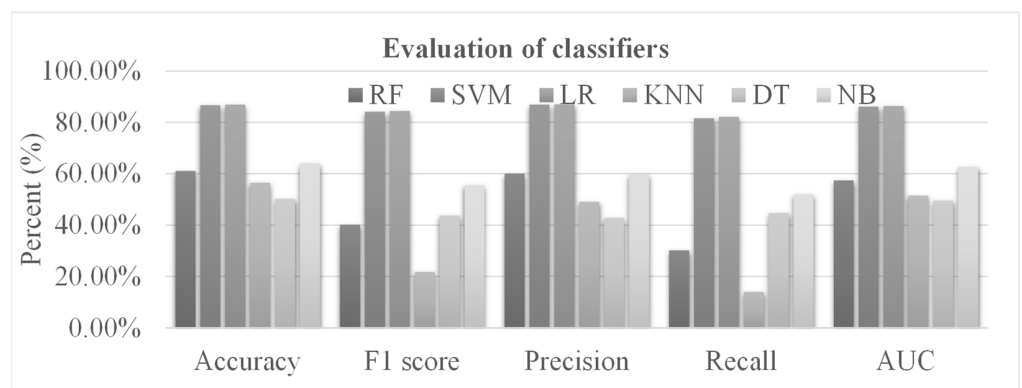
3.1. Performance of the lncRNA-Based Models

Multiple lncRNAs have been implicated in cancer onset and development [64]. Machine learning methods have been widely used to predict metastatic melanoma [11,12]. In this section, we use the lncRNA biomarkers to predict metastatic melanoma by machine learning methods and investigated which method achieved a better performance. We abbreviated the expression level of an lncRNA as an lncRNA feature in this study.

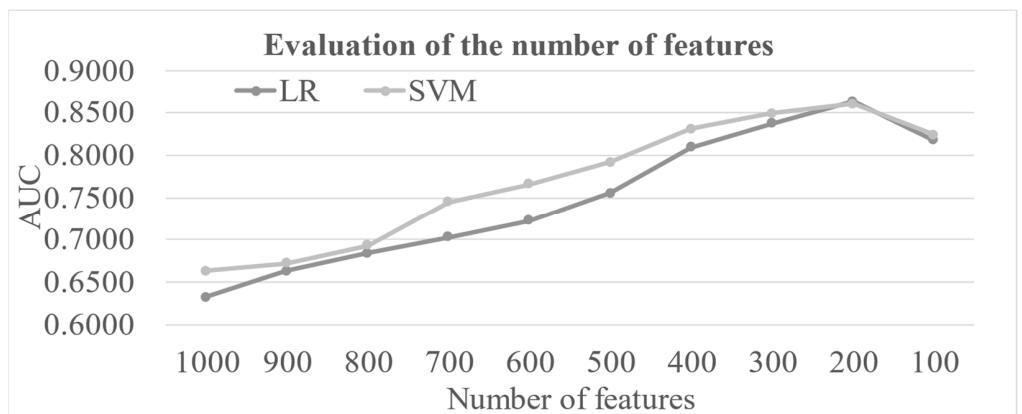
There were 6919 lncRNAs whose expression levels were profiled in the TCGA-SKCM dataset used in this study, which was much larger than that (414) of the samples. If the expression levels of all these 6919 lncRNAs (also called 6919 lncRNA features) were used to train the prediction models, it would be very easy to overfit the model and a stable prediction performance would not be achieved. We hypothesized that the feature selection methods could remove the redundant features and improve the performance of the prediction model. Therefore, feature selection algorithms were used to screen the lncRNAs whose expression levels were associated with melanoma metastasis [65,66]. Three feature selection algorithms were evaluated using lncRNA features, including SVM-RFE [49], variance [50] and t-test [51]. We used the above feature selection algorithms to select 200 lncRNAs and compared their classification performances with each other. Figure 3a shows that the model using all the features only achieved 0.5918 in accuracy. The prediction accuracy was improved to 0.8696 if SVM-RFE was used to screen the subset of metastasis-related features. SVM-RFE outperformed the t-test and variance in accuracy in selecting features for the prediction task of metastatic melanoma. The best AUC value of 0.8638 was also achieved by SVM-RFE. Therefore, the following section uses SVM-RFE as the feature selection algorithm for the lncRNA data source. Information on all the selected lncRNAs is shown in Supplementary Table S2 in the Supplementary Materials.



(a)



(b)



(c)

Figure 3. Performance evaluations of the lncRNA-based models. (a) Evaluation of the models using three feature selection algorithms and all the features (without FS). The horizontal axis gives the performance measurements. SVM was used as the classifier. Abbreviations: without FS, without the feature selection method; SVM-RFE, support vector machine recursive feature elimination. (b) Evaluation of the six classifiers using the SVM-RFE feature selection algorithm. The horizontal axis is the same as in (a). Abbreviations: RF, Random Forest; SVM, Support Vector Machine; LR, Linear Regression; KNN, K-Nearest Neighbor; DT, Decision Tree; NB, Naïve Bayes. (c) Evaluation of different numbers of features. The horizontal axis gives the number of features used in the SVM-RFE feature selection algorithm. SVM was used as the classifier. Abbreviations: SVM, support vector machine; LR, linear regression.

The classification algorithm is another important factor for prediction performance. Different models are suitable for different data types. To choose the most suitable model, this study evaluated six classifiers on the prediction task of metastatic melanoma, including random forest (RF) [57], support vector machine (SVM) [58], linear regression (LR) [59], k-nearest neighbor (KNN) [60], decision trees (DT) [61] and naïve Bayes (NB) [62]. Figure 3b shows that SVM and LR achieved the top two best prediction accuracies of 0.8671 and 0.8696, respectively. LR performed slightly better than SVM in both accuracy and AUC. Therefore, the classifier LR was used for the lncRNA data in the following sections.

The number of features selected by the feature selection algorithms was an important factor for prediction performance. The important melanoma-associated features need to be selected, but the redundant features should be eliminated from the final model. SVM and LR achieved similarly good prediction performance, shown in Figure 3b, and were further evaluated using different numbers of features, shown in Figure 3c. The AUC values increased with more features eliminated by SVM-RFE until the number of features reached 200. The AUC model decreased to 0.8178 and 0.8243 for LR and SVM, respectively. The best AUC values of 0.8638 and 0.8610 were achieved by LR and SVM using 200 features, respectively. Therefore, 200 was the default number of lncRNAs whose expression levels were chosen for the classification models.

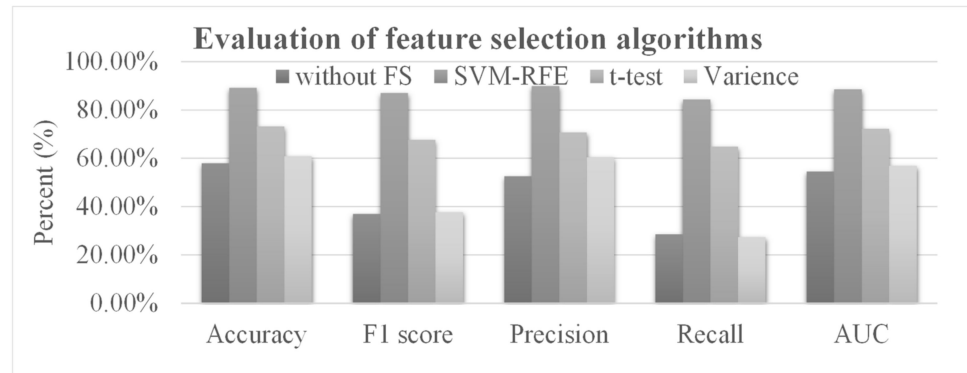
3.2. Performance of the mRNA-Based Models

Protein-coding genes represent another important component of the progression of melanoma. This section investigates how the expression levels of mRNAs (also called mRNA features) could facilitate the melanoma metastasis prediction task.

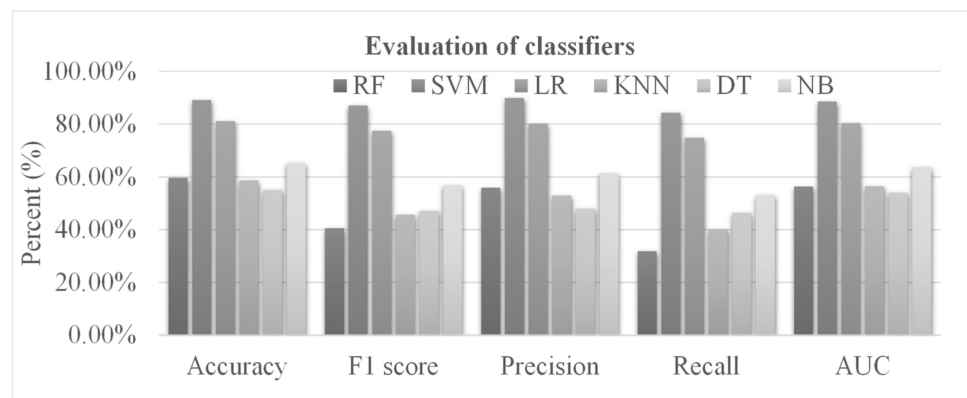
There were 19051 mRNAs whose expression levels were profiled in the TCGA-SKCM dataset. This number was also much larger than that (414) of the samples. In order to avoid the overfitting problem, feature selection algorithms were used to reduce the feature dimension. Figure 4a shows that the prediction models using all the mRNA features did not achieve a good performance and the models using the variance-selected features performed only slightly better. The features selected by SVM-RFE achieved the best accuracy of 0.8913, which was 0.3116 better in accuracy than the prediction model using all the mRNA features. Information on all the selected mRNAs is shown in Supplementary Table S3 in the Supplementary Materials.

Different classifiers can achieve different performances in a dataset. This study evaluated six classifiers on how the SVM-RFE-selected features performed on the mRNA-based metastasis prediction task, as shown in Figure 4b. The three classifiers, RF, KNN and DT, only achieved an accuracy smaller than 0.6000, while NB achieved a slightly better accuracy of 0.6522. The other two classifiers, SVM and LR, achieved the best two accuracies of 0.8913 and 0.8116, respectively. The best classifier, SVM, achieved the best AUC of 0.8856.

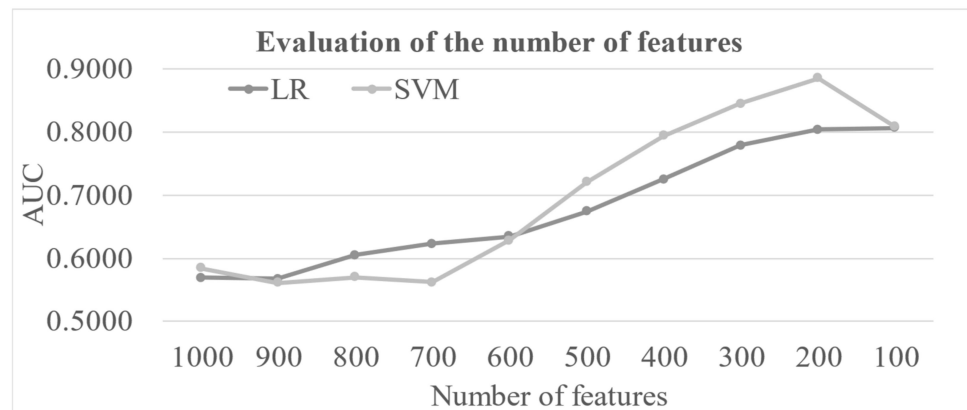
Different numbers of selected features can influence the prediction performance, as evaluated in Figure 4c. There were fluctuations in the prediction models' AUC values using 600–1000 features for both SVM and LR classifiers. After the number of features was reduced to less than 600, the prediction AUC values rapidly increased to peaks using 200 features, i.e., an AUC of 0.8041 and 0.8856 for LR and SVM, respectively. The prediction models using 100 features did not achieve the best AUC values.



(a)



(b)



(c)

Figure 4. Performance evaluations of the mRNA-based models. (a) Evaluation of the models using three feature selection algorithms and all the features (without FS). The horizontal axis gives the performance measurements. SVM was used as the classifier. Abbreviations: without FS, without feature selection method; SVM-RFE, support vector machine recursive feature elimination. (b) Evaluation of the six classifiers using the SVM-RFE feature selection algorithm. The horizontal axis is the same as in (a). Abbreviations: RF, Random Forest; SVM, Support Vector Machine; LR, Linear Regression; KNN, K-Nearest Neighbor; DT, Decision Tree; NB, Naïve Bayes. (c) Evaluation of different numbers of features. The horizontal axis gives the number of features used in the SVM-RFE feature selection algorithm. SVM was used as the classifier. Abbreviations: SVM, support vector machine; LR, linear regression.

3.3. Performance of the Image-Based Models

Pathology imaging provides an important view of cancer tissue and has been widely used in diagnosing primary and metastatic cancers [67–69]. Eight cell types were segmented and counted from the pathology images and the percentage of each cell type among all the detected cells was denoted as the image feature of this cell type for the corresponding sample. Figure 5 shows that the six classifiers used for the lncRNA and mRNA features did not achieve accuracies better than 0.6000. Therefore, we further built a fully connected neural network (MLP) for comparison with the six conventional classifiers. The MLP achieved the best accuracy of 0.6667, while the next best classifier, DT, only achieved an accuracy of 0.5821. The MLP's performance accuracy of 0.6667 was much smaller than those of the lncRNA-based and mRNA-based models. Therefore, we hypothesized that the integration of multi-modal data sources might achieve better metastasis prediction performance.

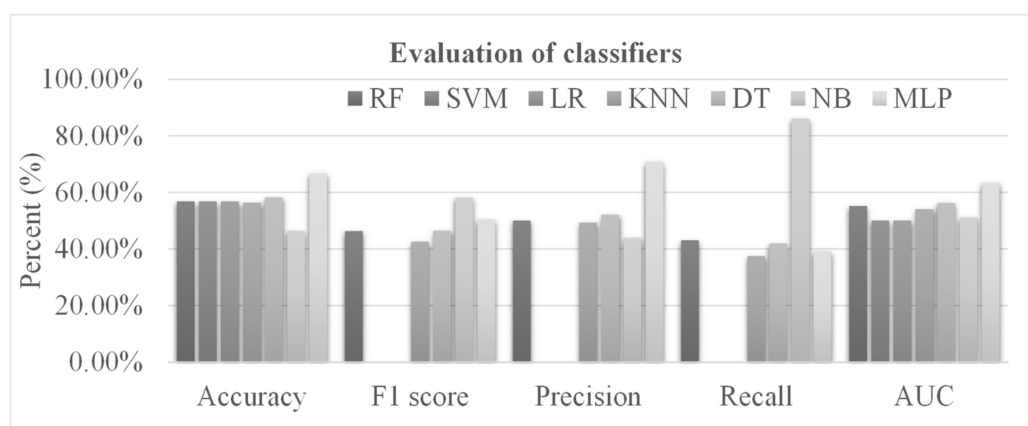


Figure 5. The performance of different classifiers using the features extracted from the pathology images. The horizontal axis lists the performance measurements. The vertical axis gives the values of these measurements. Seven classifiers were evaluated. The performance measurements of some classifiers are zero. Therefore, their corresponding bin heights are 0%, which cannot be displayed in the histogram. Abbreviations: RF, Random Forest; SVM, Support Vector Machine; LR, Linear Regression; KNN, K-Nearest Neighbor; DT, Decision Tree; NB, Naïve Bayes; MLP, Multi-Layer Perceptron.

3.4. Integration of Multimodal Data

All three data modalities (lncRNA, mRNA and image features) contributed useful information for melanoma metastasis, and we hypothesized that their integration may obtain better prediction performance. Figure 6 supports the necessity of integrating multimodal data for the metastasis prediction task. Both lncRNA and mRNA features facilitated the prediction models with accuracies >0.8500 , while the prediction model using the image-extracted features only achieved an accuracy of 0.6667. After the integration of all three data modalities, EnsembleSKCM achieved a much better prediction accuracy of 0.9444, improving the three lncRNA, mRNA and image data modalities by 0.0749, 0.0531 and 0.2778 in accuracy, respectively.

3.5. Integration of Image Features with Molecular-Level Features

Molecular-level data fully reflected the genetic information of melanoma, while image features represented the macrolevel information. To verify our ensembling hypothesis, we integrated the image features with the molecular-level features and evaluated the integration performances. As shown in Table 1, the image features may improve the model based on mRNA features by 0.0024 in accuracy and 0.0211 in AUC. The model based on lncRNA features may be improved via the integration of the image features by 0.0023 in AUC, with a slight decrease of 0.0049 in accuracy. The improved parameter-independent AUC suggests that integrating the image features provides a more balanced prediction performance. The proposed EnsembleSKCM model integrated all three data sources and

improved the model using only the mRNA and lncRNA features by 0.0024 in accuracy and 0.0030 in AUC. The above data suggest the importance of adding pathological imaging features for predicting metastatic melanoma.

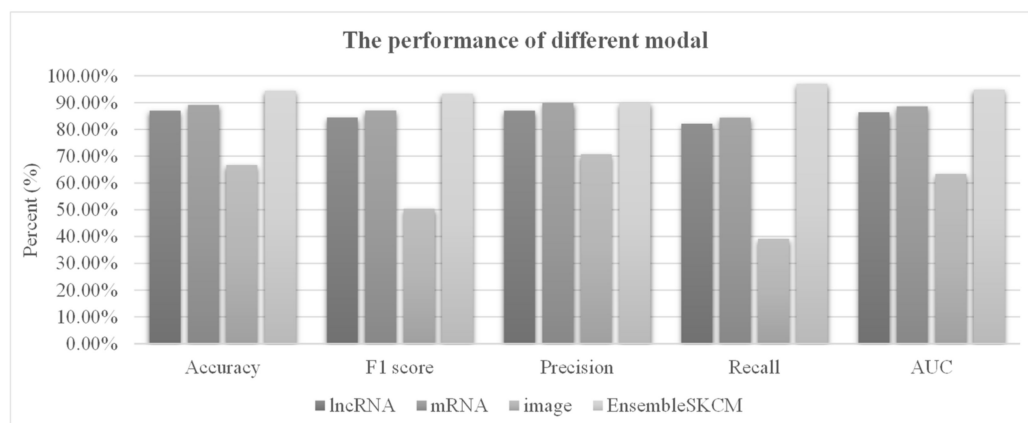


Figure 6. Contribution comparison of multimodal data for the metastasis prediction task. The horizontal axis lists the performance measurements. The vertical axis gives the values of the measurements. Three modalities were evaluated, including lncRNA, mRNA and image. lncRNA, mRNA and image features were used to train models and showed the best performance. Their integration was denoted as the proposed EnsembleSKCM model. The best model of each data modality was used. For each measurement, integrating all the data modalities can achieve the best performance.

Table 1. How the image features influence the performance of the model. mRNA + Image indicates the performance of integrating mRNA and image features. mRNA indicates the performance of mRNA features. lncRNA + Image indicates the performance of integrating lncRNA and image features. lncRNA indicates the performance of lncRNA features. lncRNA + mRNA indicates the performance of integrating lncRNA and mRNA features. EnsembleSKCM indicates the performance of all features.

	mRNA + Image	mRNA	lncRNA + Image	lncRNA	lncRNA + mRNA	Ensemble SKCM
Accuracy	0.8937	0.8913	0.8647	0.8696	0.9420	0.9444
F1 score	0.8659	0.8703	0.8372	0.8448	0.9306	0.9333
Precision	0.7933	0.8988	0.8045	0.8698	0.8994	0.8994
Recall	0.9530	0.8436	0.8727	0.8212	0.9641	0.9699
AUC	0.9067	0.8856	0.8661	0.8638	0.9456	0.9486

3.6. Correct Prediction of Samples Using Different Data Modalities

We further investigated the details of how different modalities facilitated the metastasis prediction task, as shown in Table 2. All three modalities led to metastasis prediction models with satisfying numbers of correctly predicted positive samples, i.e., primary melanoma. The lncRNA-based and mRNA-based models correctly detected approximately 0.8100 metastatic melanomas, while the image-based model correctly detected only 0.3911 metastatic melanomas. However, the image-based features represented an important view of metastatic melanoma, and its integration with the lncRNA and mRNA features improved the ensembled model to 0.8994 in the percentage of correctly predicted metastatic melanoma samples.

Table 2. The numbers of correctly predicted positive and negative samples using different modalities. The “Total” column gives the total numbers of positive and negative samples. The “lncRNA”, “mRNA” and “image” columns give the data for the individual data modalities. The last column gives the numbers of correctly predicted positive (true positives, TP) and negative (true negative, TN) samples using the multimodal EnsembleSKCM model.

	Total	lncRNA	mRNA	Image	EnsembleSKCM
TP	235	214	225	206	230
TN	179	146	145	70	161

3.7. Comparison of EnsembleSKCM with Existing Metastatic Melanoma Prediction Methods

Metastatic melanoma is a high-risk cancer, and several machine learning methods have been published to predict metastatic melanoma. Bellomo et al. used the logistic regression algorithm to combine the clinicopathologic and gene expression features to predict sentinel lymph node metastatic melanoma [10]. They achieved a prediction AUC of 0.82. Garg et al. used random forest trained with signature genes to predict metastasis and achieved the best AUC of 0.68 [11]. Mancuso et al. classified early-stage melanoma patients with high and low risk of metastasis and achieved an AUC of 0.8922 [12]. Shepelin et al. used SVM to identify 44 characteristic signaling pathways associated with melanoma metastasis [13]. Their model achieved accuracies of 0.94 for metabolic pathways and 0.923 for signaling pathways. As summarized in Table 3, the proposed EnsembleSKCM model outperformed the existing methods based on the AUC and accuracy performance metrics.

Table 3. Comparison of EnsembleSKCM with other studies. The “Publication” column gives the authors’ names and the publication date. The “Methods” column gives the method used in this study. The “Key Finding(s)” column gives the main findings of this study. The “Performance” column gives the value of evaluating indicator.

Publication	Methods	Key Finding(s)	Performance
Bellomo et al., 2020 [10]	Logistic regression model optimized by penalized maximum likelihood estimation algorithm	The model combining clinicopathologic and gene expression features better predicted SLN metastases than only one type of above features	AUC = 0.82
Garg et al., 2021 [11]	Random Forest	The machine learning models trained with signature genes performed better in predicting metastases than models trained with clinical covariates or published prognostic signatures	AUC = 0.68
Mancuso et al., 2021 [12]	Logistic Regression, Support Vector Machine, Decision Tree, Gaussian Naïve Bayes, K-Nearest Neighbors	The machine learning method that classified early-stage melanoma patients with high and low risk of metastasis by serum cytokines and Breslow thickness can best predict metastatic melanoma	AUC = 0.8922. Accuracy = 0.8502
Shepelin et al., 2018 [13]	SVM	Identified 44 characteristic signaling pathways associated with metastatic melanoma	Accuracy (metabolic pathways) = 0.94 Accuracy (signaling pathways) = 0.923
Ours	EnsembleSKCM	Integrates lncRNA, mRNA and image features to obtain better performance in recognizing metastatic melanoma	Accuracy = 0.9444. AUC = 0.9486.

3.8. Analysis of the Relationship between Molecular and Image Features Using WGCNA

To verify whether image features were redundant to molecular features, the WGCNA analysis was used to calculate the relationship between two kinds of features. As shown in Figure 7a, the soft-threshold power was defined as 3 and the scale-free topology index was

0.85, which conformed to the power law distribution. As shown in Figure 7b, when the soft threshold is 3, the curve tends to smooth and proves the good network connectivity. The gene dendrograms and respective module colors are shown in Figure 7c. We divided the molecular features into 12 modules. Figure 7d shows the relationship between the molecular modules and image features. The strongest correlation coefficient of 0.32 ($p = 5 \times 10^{-11}$) was observed between neutrophils and MEpink. We also provide information about the genes corresponding to each module in the Supplementary Table S4. There were only a few other significant correlations between the molecular modules and the image features. Therefore, most of the imaging features could contribute nonredundant complementary information to the prediction of metastatic melanoma.

3.9. Validation of the Results in Another Cohort

To further verify the validity of the model, another cohort was used to test the model. The classification models were trained using TCGA samples and tested using the GSE59455 dataset. Table 4 evaluates different classifiers using the two data sources, lncRNA and mRNA, and their integration. The classifier GBDT achieved the best accuracy of 0.6333 using the combined list of lncRNAs and mRNAs, which improved the two GBDT models using lncRNAs and mRNAs, separately. The prediction accuracies had large room for improvement due to the variations between the two transcriptomic profiling technologies' array and RNA-seq. However, the overall data support the observation that lncRNAs and mRNAs contribute complementary information to each other, and their combination leads to better prediction models.

Table 4. Accuracies of different modalities by different classifiers. The “mRNA” row gives the prediction accuracies of the mRNA-based classification models. The “lncRNA” row gives the prediction accuracies by different classifiers based on lncRNA data. The “EnsembleSKCM” row evaluates different classifiers by combining both data sources. The prediction model was trained using TCGA samples and tested on array-based transcriptomic samples from GSE59455.

	NB	GBDT	SVM	KNN	DT	LR	RF
mRNA	0.3000	0.5333	0.4667	0.4500	0.5167	0.4500	0.3667
lncRNA	0.3000	0.5333	0.2833	0.4500	0.4167	0.2833	0.4167
EnsembleSKCM	0.3000	0.6333	0.4667	0.5000	0.5167	0.4500	0.4333

The same stratified five-fold cross-validation (S5FCV) strategy was used to evaluate the proposed EnsembleSKCM algorithm on the new GSE59455 dataset, as shown in Table 5. The top 10 lncRNAs and top 10 mRNAs ranked by t-test were evaluated. The classifier NB achieved the best models on both lncRNA (Acc = 0.9500) and mRNA (accuracy = 0.8333) features, while the best prediction model (Acc = 0.9667) was achieved by combining the lncRNA and mRNA features. In summary, NB and all the other classifiers supported the importance of combining the complementary data sources of lncRNAs and mRNAs.

Table 5. Accuracies of different modalities by different classifiers. The “mRNA” row gives the prediction accuracies of the mRNA-based classification models. The “lncRNA” row gives the prediction accuracies by different classifiers based on lncRNA data. The “EnsembleSKCM” row evaluates different classifiers by combining both data sources. The prediction accuracies were calculated using the stratified five-fold cross-validation (S5FCV) strategy.

	NB	GBDT	SVM	KNN	DT	LR	RF
mRNA	0.8333	0.7167	0.7167	0.8167	0.6167	0.7167	0.8000
lncRNA	0.9500	0.8500	0.9333	0.8000	0.8500	0.9333	0.9000
EnsembleSKCM	0.9667	0.8500	0.9333	0.8500	0.8500	0.9500	0.9167

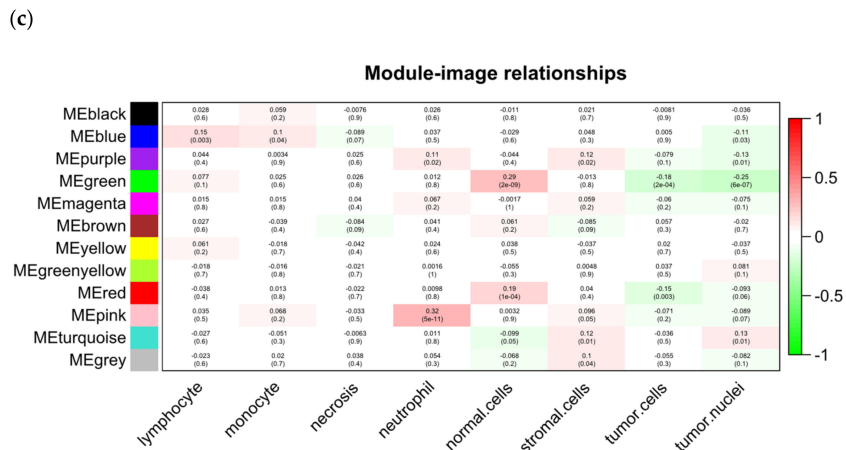
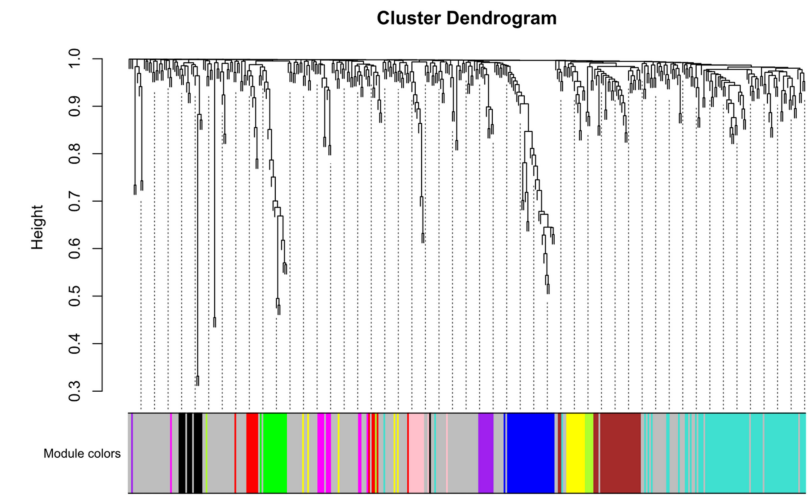
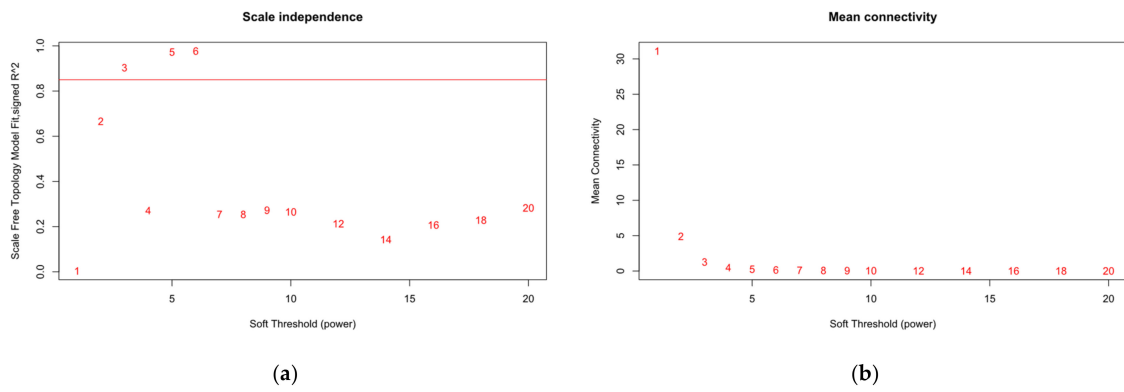


Figure 7. WGCNA analysis. (a) The network topology for different soft-threshold powers. This image shows the scale-free topology fit index influenced by soft-threshold power. The scale-free topology index was 0.85, which conformed to the power law distribution. (b) The network topology for different soft-threshold powers. This image shows mean connectivity influenced by soft-threshold power. (c) Gene clustering tree using hierarchical clustering of adjacency-based dissimilarity. (d) The module-image relationships. The correlation between the imaging features and molecular features was low; imaging features can provide complementary information to the prediction of metastatic melanoma.

4. Discussion

This study proposed the EnsembleSKCM framework to integrate the data modalities of lncRNA, mRNA and pathology images for the prediction of metastatic melanoma. The data suggest that each data modality represents an important view of the metastatic melanoma.

Some lncRNAs are known to be closely associated with the prognosis of melanoma [30,64]. Machine learning methods have already been utilized to investigate how lncRNAs are involved in the prognosis of melanoma [33]. Not all lncRNAs contributed to metastatic melanoma and the experimental design supported this through feature selection algorithms.

The view of lncRNA features alone did not achieve a satisfying prediction performance of metastatic melanoma. Therefore, the mRNA features and image-based features were also evaluated for single-modal prediction performance. The mRNA-based model achieved a similar performance as the lncRNA-based model, while the image-based model achieved a much worse performance.

However, the integration of all three data modalities generated the best model, with an accuracy rate of 0.9444. The experimental data suggest that the multimodal EnsembleSKCM model outperformed the models using only single-modal data, although the image-based model only achieved 0.3911 in the percentage of correctly predicted metastatic melanoma samples.

The lncRNA–mRNA interaction network described the close connections between the two data modalities, lncRNA and mRNA, and novel insights could be derived from the network view about melanoma compared with studies using only one modality [31,70]. This study further integrated the macrolevel image features with the molecular-level lncRNA and mRNA features. The experimental data suggest that the integration of these three data modalities may further improve the prediction performance of metastatic melanoma.

WGCNA was used to analyze the relationship between the molecular features and the image features. The data suggest that there are limited correlations between molecular and image features. Therefore, it is important to integrate both molecular and imaging features for a better prediction of metastatic melanoma.

We compared our method with the existing metastatic melanoma prediction methods. The comparison data suggest the necessity of integrating lncRNA, mRNA and image features for the prediction of metastatic melanoma. The integrated model of the three data modalities also outperformed the existing studies in this task.

Due to the limitation in data availability, the dataset used in this study is already the largest cohort. The ideal validation cohort consists of melanoma patients with paired samples before and after metastasis, and the transcriptomes are profiled by RNA-seq technology. A minimum requirement is a cohort of gender- and race-matched patients with metastatic and non-metastatic melanoma, considering the gender- [71,72] and racial disparities [73]. We only found an array-based transcriptomic dataset of metastatic melanoma to validate our method. The experimental results of both the TCGA-trained model and the proposed EnsembleSKCM algorithm support the importance of combining the complementary lncRNA and mRNA data sources. In addition to the lncRNA, mRNA and pathology image features, dermoscopic images and somatic mutations may also be considered in the integrated EnsembleSKCM framework in future studies.

This study quantitatively suggested that the precise diagnosis of metastatic melanoma may need to integrate complementary information from both molecular and macroscopic features, including lncRNAs, mRNAs and pathology images. These features represent the dynamic situations of melanoma lesions. In future research, in addition to integrating dynamic lncRNA and mRNA features, other slowly altered features, such as somatic mutations, will be evaluated in the integrated EnsembleSKCM framework for their contributions to the performance of metastatic melanoma prediction. In addition, more handcrafted feature types will be considered for the pathology images. Deep neural networks are good at automatically learning the latent patterns within images and will also be utilized to extract useful features from pathology images for the metastatic melanoma prediction task.

Additional RNA-seq transcriptomic datasets together with pathological images unbiased across multiple ethnic groups will be sought to further validate our proposed model in future studies.

5. Conclusions

This study extensively evaluated metastatic melanoma prediction models using three data modalities. The experimental data support the necessity of removing redundant features and testing different classifiers. The integration of all three data modalities also improved the single-modal models by at least 0.0531 in prediction accuracy. Metastatic melanoma has a high mortality rate, and the recently developed immunotherapy has produced major clinical success in treating lethal melanoma [74]. The precise risk assessment of melanoma provides important information for deciding follow-up treatment plans, including immunotherapy. Therefore, it is both clinically important and computationally challenging to develop precise risk assessment models for melanoma [75,76].

There are still limitations remaining in the proposed model. We trained and validated our model across different transcriptome profiling platforms (RNA-seq and microarray). Although the cross-platform validation results show that our detected transcriptome biomarkers delivered satisfactory melanoma metastasis prediction performances, the integrated model of the three data sources (lncRNA, mRNA and image) remains to be evaluated on an independent cohort. Melanoma has a 12-times higher incidence rate in the United States than in China [77]. Therefore, an independent cohort across different ethnic group will be recruited to cover the multi-modal data sources in the future studies. The experimental evaluation of the validity and robustness of our EnsembleSKCM model is also worth future studies for the detection of metastatic melanoma in the clinical practice.

System biology models have the capability of integrating heterogeneous data sources in network settings. We plan to explore the possibility of combining systems biology and machine learning approaches via the graph convolutional network for the prediction of melanoma metastasis.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13101916/s1>, Table S1. The metadata information of the TCGA dataset. The sample names were in the column "Sample". The classes of "Primary" or "Metastasis" were in the column "Class". The other metadata are listed in the other columns. Table S2. Summary of lncRNA selected by the feature selection algorithm. The ensembl IDs were given in the column "ENSEMBL", the chromosome names were given in the column "Chr", the start coordinates of genes were given in the column "Start", the end coordinates of genes were given in the column "End", the gene names were given in the column "SYMBOL". Table S3. Summary of mRNA selected by the feature selection algorithm. The ensembl IDs were given in the column "ENSEMBL", the chromosome names were given in the column "Chromosome", the start coordinates of genes were given in the column "Start", the end coordinates of genes were given in the column "End", the gene names were given in the column "SYMBOL". Table S4. The information about the gene corresponding to each WGCNA module. The ensembl IDs were given in the column "Gene Name", the module names were given in the column "WGCNA-Module".

Author Contributions: F.Z. and M.D. conceived and designed this study; S.L., Y.F. and K.L. wrote the program and carried out the experiments; S.L., H.Z., X.W. and R.J. designed and carried out the revision experiments and data analysis; S.L., M.D. and L.H. drafted the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Senior and Junior Technological Innovation Team (20210509055RQ), the National Natural Science Foundation of China (62072212 and U19A2061), the Jilin Provincial Key Laboratory of Big Data Intelligent Computing (20180622002JC), and the Fundamental Research Funds for the Central Universities, JLU.

Acknowledgments: Evaluation and discussion of the validity of the proposed EnsembleSKCM model and the future experimental design from Qi Qi (First People’s Hospital of Yancheng) and Jiannan Huang (Jilin Cancer Hospital) are much appreciated. The insightful comments from the two anonymous reviewers improved this manuscript a lot and are greatly appreciated!

Conflicts of Interest: The authors declare no conflict of interest.

References

- Schadendorf, D.; Fisher, D.E.; Garbe, C.; Gershenwald, J.E.; Grob, J.-J.; Halpern, A.; Herlyn, M.; Marchetti, M.A.; McArthur, G.; Ribas, A. Melanoma. *Nat. Rev. Dis. Prim.* **2015**, *1*, 1–20. [[CrossRef](#)] [[PubMed](#)]
- Dimitriou, F.; Krattinger, R.; Rameyte, E.; Barysch, M.J.; Micalletto, S.; Dummer, R.; Goldinger, S.M. The world of melanoma: Epidemiologic, genetic and anatomic differences of melanoma across the globe. *Curr. Oncol. Rep.* **2018**, *20*, 87. [[CrossRef](#)] [[PubMed](#)]
- Guhan, S.; Boland, G.; Tanabe, K.; Lin, W.; Reddy, B.; Hawryluk, E.B.; Sober, A.J.; Tsao, H. Surgical delay and mortality for primary cutaneous melanoma. *J. Am. Acad. Dermatol.* **2021**, *84*, 1089–1091. [[CrossRef](#)] [[PubMed](#)]
- Ghazawi, F.M.; Cyr, J.; Darwich, R.; Le, M.; Rahme, E.; Moreau, L.; Netchiporouk, E.; Zubarev, A.; Roshdy, O.; Glassman, S.J. Cutaneous malignant melanoma incidence and mortality trends in Canada: A comprehensive population-based study. *J. Am. Acad. Dermatol.* **2019**, *80*, 448–459. [[CrossRef](#)]
- Berk-Krauss, J.; Stein, J.A.; Weber, J.; Polsky, D.; Geller, A.C. New systematic therapies and trends in cutaneous melanoma deaths among US whites, 1986–2016. *Am. J. Public Health* **2020**, *110*, 731–733. [[CrossRef](#)] [[PubMed](#)]
- Cortez, J.L.; Vasquez, J.; Wei, M.L. The impact of demographics, socioeconomics, and health care access on melanoma outcomes. *J. Am. Acad. Dermatol.* **2021**, *84*, 1677–1683. [[CrossRef](#)] [[PubMed](#)]
- Verma, B.K.; Subramaniam, P.; Vadigepalli, R. Model-based virtual patient analysis of human liver regeneration predicts critical perioperative factors controlling the dynamic mode of response to resection. *BMC Syst. Biol.* **2019**, *13*, 9. [[CrossRef](#)]
- Verma, B.K.; Subramaniam, P.; Vadigepalli, R. Characterizing different class of patients based on their liver regeneration capacity post hepatectomy and the prediction of safe future liver volume for improved recovery. In Proceedings of the 2018 International Conference on Bioinformatics and Systems Biology, Las Vegas, NV, USA, 19–21 March 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 152–156.
- Milberg, O.; Gong, C.; Jafarnejad, M.; Bartelink, I.H.; Wang, B.; Vicini, P.; Narwal, R.; Roskos, L.; Popel, A.S. A QSP model for predicting clinical responses to monotherapy, combination and sequential therapy following CTLA-4, PD-1, and PD-L1 checkpoint blockade. *Sci. Rep.* **2019**, *9*, 11286. [[CrossRef](#)] [[PubMed](#)]
- Bellomo, D.; Arias-Mejias, S.M.; Ramana, C.; Heim, J.B.; Quattrocchi, E.; Sominidi-Damodaran, S.; Bridges, A.G.; Lehman, J.S.; Hieken, T.J.; Jakub, J.W. Model combining tumor molecular and clinicopathologic risk factors predicts sentinel lymph node metastasis in primary cutaneous melanoma. *JCO Precis. Oncol.* **2020**, *4*, 319–334. [[CrossRef](#)] [[PubMed](#)]
- Garg, M.; Couturier, D.-L.; Nsengimana, J.; Fonseca, N.A.; Wongchenko, M.; Yan, Y.; Lauss, M.; Jönsson, G.B.; Newton-Bishop, J.; Parkinson, C. Tumour gene expression signature in primary melanoma predicts long-term outcomes. *Nat. Commun.* **2021**, *12*, 1137. [[CrossRef](#)]
- Mancuso, F.; Lage, S.; Rasero, J.; Díaz-Ramón, J.L.; Apraiz, A.; Pérez-Yarza, G.; Ezkurra, P.A.; Penas, C.; Sánchez-Diez, A.; García-Vazquez, M.D. Serum markers improve current prediction of metastasis development in early-stage melanoma patients: A machine learning-based study. *Mol. Oncol.* **2020**, *14*, 1705–1718. [[CrossRef](#)] [[PubMed](#)]
- Shepelin, D.; Korzinkin, M.; Vanyushina, A.; Aliper, A.; Borisov, N.; Vasilov, R.; Zhukov, N.; Sokov, D.; Prassolov, V.; Gaifullin, N. Molecular pathway activation features linked with transition from normal skin to primary and metastatic melanomas in human. *Oncotarget* **2016**, *7*, 656. [[CrossRef](#)] [[PubMed](#)]
- Tsao, H.; Olazagasti, J.M.; Cordoro, K.M.; Brewer, J.D.; Taylor, S.C.; Bordeaux, J.S.; Chren, M.-M.; Sober, A.J.; Tegeler, C.; Bhushan, R.; et al. Early detection of melanoma: Reviewing the ABCDEs. *J. Am. Acad. Dermatol.* **2015**, *72*, 717–723. [[CrossRef](#)] [[PubMed](#)]
- Brinker, T.J.; Hekler, A.; Enk, A.H.; Berking, C.; Haferkamp, S.; Hauschild, A.; Weichenthal, M.; Klode, J.; Schadendorf, D.; Holland-Letz, T. Deep neural networks are superior to dermatologists in melanoma image classification. *Eur. J. Cancer* **2019**, *119*, 11–17. [[CrossRef](#)]
- Hekler, A.; Utikal, J.S.; Enk, A.H.; Berking, C.; Klode, J.; Schadendorf, D.; Jansen, P.; Franklin, C.; Holland-Letz, T.; Krahl, D. Pathologist-level classification of histopathological melanoma images with deep neural networks. *Eur. J. Cancer* **2019**, *115*, 79–83. [[CrossRef](#)]
- Mo, Q.; Wan, L.; Schell, M.J.; Jim, H.; Tworoger, S.S.; Peng, G. Integrative Analysis Identifies Multi-Omics Signatures That Drive Molecular Classification of Uveal Melanoma. *Cancers* **2021**, *13*, 6168. [[CrossRef](#)]
- Gadeyne, L.; Van Herck, Y.; Milli, G.; Atak, Z.K.; Bolognesi, M.M.; Wouters, J.; Marcelis, L.; Minia, A.; Pliaka, V.; Roznac, J. A Multi-Omics Analysis of Metastatic Melanoma Identifies a Germinal Center-Like Tumor Microenvironment in HLA-DR-Positive Tumor Areas. *Front. Oncol.* **2021**, *11*, 787. [[CrossRef](#)]
- Khamparia, A.; Singh, P.K.; Rani, P.; Samanta, D.; Khanna, A.; Bhushan, B. An internet of health things-driven deep learning framework for detection and classification of skin cancer using transfer learning. *Trans. Emerg. Telecommun. Technol.* **2021**, *32*, e3963. [[CrossRef](#)]

20. Thomas, S.M.; Lefevre, J.G.; Baxter, G.; Hamilton, N.A. Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer. *Med. Image Anal.* **2021**, *68*, 101915. [[CrossRef](#)]
21. Duggani, K.; Nath, M.K. A Technical Review Report on Deep Learning Approach for Skin Cancer Detection and Segmentation. *Data Anal. Manag.* **2021**, 87–99.
22. Widiensyah, M.; Rasyid, S.; Wisnu, P.; Wibowo, A. Image segmentation of skin cancer using MobileNet as an encoder and linknet as a decoder. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2021; p. 012113.
23. Mohakud, R.; Dash, R. Skin cancer image segmentation utilizing a novel EN-GWO based hyper-Parameter optimized FCEDN. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**. [[CrossRef](#)]
24. Tang, Y.; Yang, F.; Yuan, S. A multi-Stage framework with context information fusion structure for skin lesion segmentation. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging, Prague, Czech Republic, 13–16 April 2016; IEEE: Piscataway, NJ, USA, 2019; pp. 1407–1410.
25. Saleem, A.; Bhatti, M.N.A.; Ashraf, M.A.; Zia, M.; Mahmood, H. Segmentation and classification of consumer-grade and dermoscopic skin cancer images using hybrid textural analysis. *J. Med. Imaging* **2019**, *6*, 034501.
26. Lai, X.; Zhou, J.; Wessely, A.; Heppt, M.; Maier, A.; Berking, C.; Vera, J.; Zhang, L. A disease network-based deep learning approach for characterizing melanoma. *Int. J. Cancer* **2021**, *150*, 1029–1044. [[CrossRef](#)] [[PubMed](#)]
27. Wei, D. A multigene support vector machine predictor for metastasis of cutaneous melanoma. *Mol. Med. Rep.* **2018**, *17*, 2907–2914. [[CrossRef](#)] [[PubMed](#)]
28. Korfiati, A.; Livanos, G.; Konstantinou, C.; Georgiou, S.; Sakellaropoulos, G. eBioMelDB: Multi-Modal Database for Melanoma and Its Application on Estimating Patient Prognosis. In Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations., Suzhou, China, 15–17 October 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 33–44.
29. Liu, N.; Liu, Z.; Liu, X.; Chen, H. Comprehensive analysis of a competing endogenous RNA network identifies seven-lncRNA signature as a prognostic biomarker for melanoma. *Front. Oncol.* **2019**, *9*, 935. [[CrossRef](#)]
30. Xia, Y.; Zhou, Y.; Han, H.; Li, P.; Wei, W.; Lin, N. lncRNA NEAT1 facilitates melanoma cell proliferation, migration, and invasion via regulating miR-495-3p and E2F3. *J. Cell. Physiol.* **2019**, *234*, 19592–19601. [[CrossRef](#)]
31. Zhang, J.; Liu, H.; Zhang, W.; Li, Y.; Fan, Z.; Jiang, H.; Luo, J. Identification of lncRNA-mRNA regulatory module to explore the pathogenesis and prognosis of melanoma. *Front. Cell Dev. Biol.* **2020**, *8*, 1584. [[CrossRef](#)]
32. Schmidt, K.; Carroll, J.S.; Yee, E.; Thomas, D.D.; Wert-Lamas, L.; Neier, S.C.; Sheynkman, G.; Ritz, J.; Novina, C.D. The lncRNA SLNCR recruits the androgen receptor to EGR1-bound genes in melanoma and inhibits expression of tumor suppressor p21. *Cell Rep.* **2019**, *27*, 2493–2507. [[CrossRef](#)]
33. Yan, K.; Wang, Y.; Shao, Y.; Xiao, T. Gene Instability-Related lncRNA Prognostic Model of Melanoma Patients via Machine Learning Strategy. *J. Oncol.* **2021**, *2021*, 5582920. [[CrossRef](#)]
34. Siena, A.D.D.; Barros, I.; Storti, C.B.; de Biagi Junior, C.A.O.; da Costa Carvalho, L.A.; Maria-Engler, S.S.; Sousa, J.F.; Silva, W.A., Jr. Upregulation of the novel lncRNA U731166 is associated with migration, invasion and vemurafenib resistance in melanoma. *J. Cell Mol. Med.* **2022**, *26*, 671–683. [[CrossRef](#)]
35. Gao, J.; Li, P.; Chen, Z.; Zhang, J. A survey on deep learning for multimodal data fusion. *Neural Comput.* **2020**, *32*, 829–864. [[CrossRef](#)] [[PubMed](#)]
36. Lahat, D.; Adali, T.; Jutten, C. Multimodal data fusion: An overview of methods, challenges, and prospects. *Proc. IEEE* **2015**, *103*, 1449–1477. [[CrossRef](#)]
37. Clavel, J.; Aristide, L.; Morlon, H. A Penalized Likelihood Framework for High-Dimensional Phylogenetic Comparative Methods and an Application to New-World Monkeys Brain Evolution. *Syst Biol* **2019**, *68*, 93–116. [[CrossRef](#)] [[PubMed](#)]
38. Cai, J.; Luo, J.; Wang, S.; Yang, S. Feature selection in machine learning: A new perspective. *Neurocomputing* **2018**, *300*, 70–79. [[CrossRef](#)]
39. Khaire, U.M.; Dhanalakshmi, R. Stability of feature selection algorithm: A review. *J. King Saud Univ. Comput. Inf. Sci.* **2019**, *34*, 1060–1073. [[CrossRef](#)]
40. Urbanowicz, R.J.; Meeker, M.; La Cava, W.; Olson, R.S.; Moore, J.H. Relief-Based feature selection: Introduction and review. *J. Biomed. Inform.* **2018**, *85*, 189–203. [[CrossRef](#)]
41. Wang, X.; Li, G.; Luo, Q.; Xie, J.; Gan, C. Integrated TCGA analysis implicates lncRNA CTB-193M12. 5 as a prognostic factor in lung adenocarcinoma. *Cancer Cell Int.* **2018**, *18*, 27. [[CrossRef](#)]
42. Wu, M.; Shang, X.; Sun, Y.; Wu, J.; Liu, G. Integrated analysis of lymphocyte infiltration-associated lncRNA for ovarian cancer via TCGA, GTEx and GEO datasets. *PeerJ* **2020**, *8*, e8961. [[CrossRef](#)]
43. Akbani, R.; Akdemir, K.C.; Aksoy, B.A.; Albert, M.; Ally, A.; Amin, S.B.; Arachchi, H.; Arora, A.; Auman, J.T.; Ayala, B. Genomic classification of cutaneous melanoma. *Cell* **2015**, *161*, 1681–1696. [[CrossRef](#)]
44. Carpenter, A.E.; Jones, T.R.; Lamprecht, M.R.; Clarke, C.; Kang, I.H.; Friman, O.; Guertin, D.A.; Chang, J.H.; Lindquist, R.A.; Moffat, J. CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **2006**, *7*, R100. [[CrossRef](#)]
45. Walts, A.E.; Mirocha, J.M.; Marchevsky, A.M. Challenges in Ki-67 assessments in pulmonary large-cell neuroendocrine carcinomas. *Histopathology* **2021**, *78*, 699–709. [[CrossRef](#)]
46. Meyer, H.; Reudenbach, C.; Hengl, T.; Katurji, M.; Nauss, T. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environ. Model. Softw.* **2018**, *101*, 1–9. [[CrossRef](#)]

47. Ying, X. An overview of overfitting and its solutions. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2019; p. 022022.
48. Venkatesh, B.; Anuradha, J. A review of feature selection and its methods. *Cybern. Inf. Technol.* **2019**, *19*, 3–26. [[CrossRef](#)]
49. Albashish, D.; Hammouri, A.I.; Braik, M.; Atwan, J.; Sahran, S. Binary biogeography-based optimization based SVM-RFE for feature selection. *Appl. Soft Comput.* **2021**, *101*, 107026. [[CrossRef](#)]
50. Kamalov, F.; Moussa, S.; El Khatib, Z.; Mnaouer, A.B. Orthogonal Variance-Based Feature Selection for Intrusion Detection Systems. 2021, pp. 1–5. Available online: <https://www.sciencedirect.com/science/article/abs/pii/S092552732030205X> (accessed on 4 September 2022).
51. Ramaswamy, R.; Kandhasamy, P.; Palaniswamy, S. Feature Selection for Alzheimer’s Gene Expression Data Using Modified Binary Particle Swarm Optimization. *IETE J. Res.* **2021**, 1–12. [[CrossRef](#)]
52. Zeng, N.; Qiu, H.; Wang, Z.; Liu, W.; Zhang, H.; Li, Y. A new switching-delayed-PSO-Based optimized SVM algorithm for diagnosis of Alzheimer’s disease. *Neurocomputing* **2018**, *320*, 195–202. [[CrossRef](#)]
53. Jahed Armaghani, D.; Asteris, P.G.; Askarian, B.; Hasanipanah, M.; Tarinejad, R.; Huynh, V.V. Examining hybrid and single SVM models with different kernels to predict rock brittleness. *Sustainability* **2020**, *12*, 2229. [[CrossRef](#)]
54. Sanz, H.; Valim, C.; Vegas, E.; Oller, J.M.; Reverter, F. SVM-RFE: Selection and visualization of the most relevant features through non-Linear kernels. *BMC Bioinform.* **2018**, *19*, 432. [[CrossRef](#)]
55. Xue, Y.; Zhang, L.; Wang, B.; Zhang, Z.; Li, F. Nonlinear feature selection using Gaussian kernel SVM-RFE for fault diagnosis. *Appl. Intell.* **2018**, *48*, 3306–3331. [[CrossRef](#)]
56. Gao, S.; Wang, P.; Feng, Y.; Xie, X.; Duan, M.; Fan, Y.; Liu, S.; Huang, L.; Zhou, F. RIFS2D: A two-dimensional version of a randomly restarted incremental feature selection algorithm with an application for detecting low-ranked biomarkers. *Comput. Biol. Med.* **2021**, *133*, 104405. [[CrossRef](#)]
57. Chen, Y.; Zheng, W.; Li, W.; Huang, Y. Large group activity security risk assessment and risk early warning based on random forest algorithm. *Pattern Recognit. Lett.* **2021**, *144*, 1–5. [[CrossRef](#)]
58. Zhou, J.; Huang, S.; Wang, M.; Qiu, Y. Performance evaluation of hybrid GA–SVM and GWO–SVM models to predict earthquake-induced liquefaction potential of soil: A multi-Dataset investigation. *Eng. Comput.* **2021**, 1–19. [[CrossRef](#)]
59. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. Linear regression. In *An Introduction to Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 59–128.
60. Kück, M.; Freitag, M. Forecasting of customer demands for production planning by local k-nearest neighbor models. *Int. J. Prod. Econ.* **2021**, *231*, 107837. [[CrossRef](#)]
61. Goswami, S.; Pramanick, R.; Patra, A.; Rath, S.P.; Foltin, M.; Ariando, A.; Thompson, D.; Venkatesan, T.; Goswami, S.; Williams, R.S. Decision trees within a molecular memristor. *Nature* **2021**, *597*, 51–56. [[CrossRef](#)] [[PubMed](#)]
62. Zhang, H.; Jiang, L.; Yu, L. Attribute and instance weighted naive Bayes. *Pattern Recognit.* **2021**, *111*, 107674. [[CrossRef](#)]
63. Wang, M.; Wang, L.; Pu, L.; Li, K.; Feng, T.; Zheng, P.; Li, S.; Sun, M.; Yao, Y.; Jin, L. LncRNAs related key pathways and genes in ischemic stroke by weighted gene co-Expression network analysis (WGCNA). *Genomics* **2020**, *112*, 2302–2308. [[CrossRef](#)]
64. Yang, G.; Lu, X.; Yuan, L. LncRNA: A link between RNA and cancer. *Biochim. Et Biophys. Acta (BBA)-Gene Regul. Mech.* **2014**, *1839*, 1097–1109. [[CrossRef](#)]
65. Agrawal, P.; Abutarboush, H.F.; Ganesh, T.; Mohamed, A.W. Metaheuristic algorithms on feature selection: A survey of one decade of research (2009–2019). *IEEE Access* **2021**, *9*, 26766–26791. [[CrossRef](#)]
66. Dhiman, G.; Oliva, D.; Kaur, A.; Singh, K.K.; Vimal, S.; Sharma, A.; Cengiz, K. BEPO: A novel binary emperor penguin optimizer for automatic feature selection. *Knowl.-Based Syst.* **2021**, *211*, 106560. [[CrossRef](#)]
67. Lu, C.; Mandal, M. Automated analysis and diagnosis of skin melanoma on whole slide histopathological images. *Pattern Recognit.* **2015**, *48*, 2738–2750. [[CrossRef](#)]
68. Van Zon, M.; Stathonikos, N.; Blokk, W.A.M.; Komina, S.; Maas, S.L.N.; Pluim, J.P.W.; Van Diest, P.J.; Veta, M. Segmentation and classification of melanoma and nevus in whole slide images. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging, Iowa City, IA, USA, 3–7 April 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 263–266.
69. Peng, Y.; Chu, Y.; Chen, Z.; Zhou, W.; Wan, S.; Xiao, Y.; Zhang, Y.; Li, J. Combining texture features of whole slide images improves prognostic prediction of recurrence-free survival for cutaneous melanoma patients. *World J. Surg. Oncol.* **2020**, *18*, 130. [[CrossRef](#)] [[PubMed](#)]
70. Zhu, J.; Deng, J.; Zhang, L.; Zhao, J.; Zhou, F.; Liu, N.; Cai, R.; Wu, J.; Shu, B.; Qi, S. Reconstruction of lncRNA-miRNA-mRNA network based on competitive endogenous RNA reveals functional lncRNAs in skin cutaneous melanoma. *BMC Cancer* **2020**, *20*, 1–20. [[CrossRef](#)] [[PubMed](#)]
71. Buxeda, A.; Redondo-Pachon, D.; Perez-Saez, M.J.; Crespo, M.; Pascual, J. Sex differences in cancer risk and outcomes after kidney transplantation. *Transpl. Rev.* **2021**, *35*, 100625. [[CrossRef](#)] [[PubMed](#)]
72. Davis, D.S.; Robinson, C.; Callender, V.D. Skin cancer in women of color: Epidemiology, pathogenesis and clinical manifestations. *Int. J. Womens Derm.* **2021**, *7*, 127–134. [[CrossRef](#)] [[PubMed](#)]
73. Cooper, R.M.; Chung, J.; Hogan, T.; Haque, R. Patterns of overall mortality by race/ethnicity and socioeconomic status in insured cancer patients in Southern California. *Cancer Causes Control* **2021**, *32*, 609–616. [[CrossRef](#)] [[PubMed](#)]
74. Herzberg, B.; Fisher, D.E. Metastatic melanoma and immunotherapy. *Clin. Immunol.* **2016**, *172*, 105–110. [[CrossRef](#)] [[PubMed](#)]

75. Hersey, P.; Coates, A.S.; McCarthy, W.H.; Thompson, J.F.; Sillar, R.W.; McLeod, R.; Gill, P.G.; Coventry, B.J.; McMullen, A.; Dillon, H. Adjuvant immunotherapy of patients with high-risk melanoma using vaccinia viral lysates of melanoma: Results of a randomized trial. *J. Clin. Oncol.* **2002**, *20*, 4181–4190. [[CrossRef](#)]
76. Ma, E.Z.; Hoegler, K.M.; Zhou, A.E. Bioinformatic and Machine Learning Applications in Melanoma Risk Assessment and Prognosis: A Literature Review. *Genes* **2021**, *12*, 1751. [[CrossRef](#)]
77. Xia, C.; Dong, X.; Li, H.; Cao, M.; Sun, D.; He, S.; Yang, F.; Yan, X.; Zhang, S.; Li, N.; et al. Cancer statistics in China and United States, 2022: Profiles, trends, and determinants. *Chin. Med. J.* **2022**, *135*, 584–590. [[CrossRef](#)] [[PubMed](#)]