Istituto Tumori "Giovanni Paolo II"
IRCCS Bari

Recurrence-free survival prediction in melanoma patients by exploiting artificial intelligence techniques on melanoma whole slide images

Sabino Strippoli, Maria Colomba Comes, Samantha Bove, Ivana De Risi, Annarita Fanizzi, Livia Fucci, Fabio Mele, Alfredo Zito, Raffaella Massafra, Michele Guida

[1] I.R.C.C.S. Istituto Tumori "Giovanni Paolo II", Viale Orazio Flacco 65, 70124 Bari, Italy

Introduction

Among the melanoma cases, over 90% is represented from cutaneous melanoma, which is one of the most aggressive forms of skin cancer with a high mortality rate [1]. A prognosis improvement in cutaneous melanoma patients is crucial to better plan personalized treatments. Although more and more advanced treatments for melanoma have constantly introduced in clinical practice over the years, they can cause toxicity and overtreatment, especially for early-stage patients. Currently, clinical prognosis methods for the evaluation of the risk of recurrence includes multiple parameters at the basis of the American Joint Committee on Cancer (AJCC) pathologic tumor stage [2]. Despite routinely applied in clinical practice, these clinical prognosis methods have some pitfalls. Among them, the evaluation of complete staging is performed by means of lymph node examination, that is a matter of an ongoing scientific debate due to the associated significant post-operative morbidity and/or infection. Meanwhile, genomic-based tools complementing the traditional staging system are being developed in order to evaluate their prognostic power in comparison with traditional factors. However, these tools are currently in the experimentation phase and have not yet been applied in actual clinical practice. Thus, finding more reliable and widely applicable prognostic biomarkers in melanoma patient is urgent. Within this emerging scenario, digital pathology image-based prediction models can be designed. Due to ongoing developments in technology, e.g., cloud storage systems and computer processing powers, whole slide images (WSIs), which refers to digital slides, have become the predominant imaging modality in pathology departments across the world.

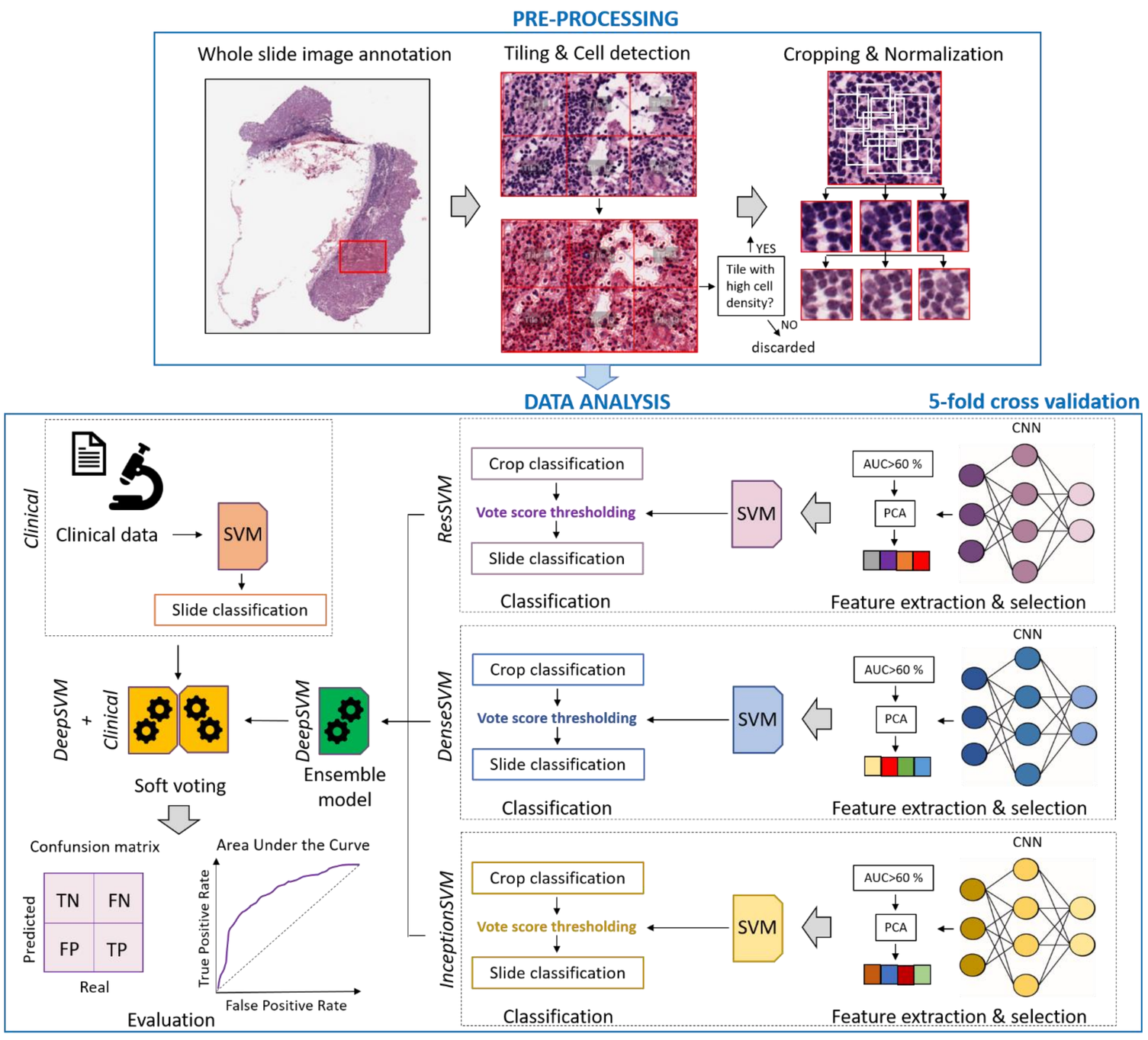

Fig. 1. Workflow of the proposed approach.

Materials and Methods

In this study, we propose a deep learning model which makes use of features extracted by transfer learning to predict 1-year disease-free survival in patients with cutaneous melanoma. A binary classification task was developed to give the prediction of 1-year disease-free survival in melanoma patients starting from the analysis of whole slide images (WSIs) representing cutaneous melanoma. Hence, patients were discerned in disease free (DF) and non-disease-free (non-DF) cases. Whole slide images referred to a cohort of 43 patients from Clinical Proteomic Tumor Analysis Consortium Cutaneous Melanoma (CPTAC-CM) public database were firstly analyzed to design the predictive model (Table 1). Then, the model was validated on independent test, i.e., a validation cohort of 11 cutaneous melanoma patients referred to our Institute (Table 2). Image pre-processing was firstly performed: WSIs were annotated by expert pathologists; the identified Region of Interests (ROIs), one per WSI, were tessellated into tiles of 224 × 224 pixels; cell detection was performed and only tiles with high cell density were retained and then divided into low-dimensional crops, which were finally colour-normalized. Data-analysis was then performed: crops were taken in input by three pre-trained CNNs. Each CNN extracted thousands of imaging features, later undergoing to a feature selection process. In correspondence of the three CNNs, three models, called *ResSVM*, *DenseSVM* and *InceptionSVM* were defined. An ensemble model, named *DeepSVM*, was designed by combining the classification scores of the three models. A model, named *Clinical*, which took in input clinical data and used a SVM classifier to give in output classification scores at WSI level, was defined. A soft-voting procedure was implemented to combine the classification scores of *DeepSVM* and *Clinical* at WSI level and then at patient level. The final model was called *DeepSVM + Clinical.* Predictive performances were assessed by standard evaluation metrics (Fig. 1).

Characteristic	Distribution
Outcome	
DF cases (%)	31 (72.1%)
non-DF (%)	12 (27.9%)
Gender	
Male (%)	21 (48.8%)
Female (%)	22 (51.2%)
Tumor site	
Trunk (%)	20 (46.6%)
Palms and (%)	1 (2.3%)
Extremities (abs.; %)	11 (25.6%)
Head & neck (%)	1 (2.3%)
Lymph nodes (%)	5 (11.6%)
Other (%)	5 (11.6%)
Stage	
I (abs.; %)	6 (14.0%)
II (abs.; %)	27 (62.7%)
III (abs.; %)	10 (23.3%)
T	
T0 (abs.; %)	2 (4.7%)
T1 (abs.; %)	3 (6.8%)
T2 (abs.; %)	10 (23.3%)
T3 (abs.; %)	6 (14.0%)
T4 (abs.; %)	22 (51.2%)
Age	
Median (std.)	64.0 (14.9)

Table 1. Clinical data referred to CPTAC-CM public dataset.

Characteristic	Distribution
Outcome	
DF cases (%)	8 (72.7%)
non-DF (%)	3 (27.3%)
Gender	
Male (%)	3 (27.3%)
Female (%)	8 (72.7%)
Tumor site	
Trunk (%)	5 (45.5%)
Extremities (%)	3 27.3%)
Head & neck (%)	1 (9.1%)
Other (%)	2 (18.2%)
Stage	
I (%)	2 (18.2%)
II (%)	9 (81.8%)
T	
T2 (%)	2 (18.2%)
T3 (%)	4 (36.3%)
T4 (%)	5 (45.5%)
Age	
Median (std.)	56.0 (13.2)

Table 2. Clinical data referred to the validation cohort of patients.

Results

With respect to the *Clinical* model, *DeepSVM* model led to great improvement for all the performance evaluation metrics: a median AUC value of 69.5%, a median accuracy value of 72.7%, a median sensitivity value of 68.8% and a median specificity value of 75.0% were obtained, respectively. The sensitivity and specificity values were well balanced. The *DeepSVM + clinical* model, defined by combining the scores achieved by *DeepSVM* and *Clinical* separately via a soft voting technique, did not lead to an overall improvement of the results with respect to the *DeepSVM* model. The only metric which increased was the median specificity value (83.3%), but at the expense of the median sensitivity value (59.4%). However, if compared with the *Clinical* model alone, the *DeepSVM + clinical* model returned better performance for all the evaluation metrics. The model which showed the best performances was those obtained by exploiting the image information alone, i.e., *DeepSVM*. This best performing model was finally tested on the validation cohort of patients recruited from our Institute, reaching an AUC value of 66.7%, an accuracy value of 72.7%, a sensitivity value of 100%, and a specificity value of 62.0%. The achieved results demonstrated how the proposed model was quite robust and generalizable.

Conclusions

A fundamental peculiarity of the proposed model is the automatic identification of quantitative imaging information from the raw WSIs directly. In other words, we used a computerized system to automatically extract information that are usually evaluated manually and visually by pathologists. Our model was able to automatically capture fine tumor or lymphocytic infiltrate characteristics, such as morphology of tumor nuclei as well as density distribution of lymphocytes, that are well-known to be associated with metastasis and survival outcomes. The promising results achieved in this preliminary work suggest how our proposal, after further validations of wider cohorts of patients as well as technical refinement, has the potential to fulfil the predictive task with great improvement in the melanoma patient management in terms of time and costs, also representing a complementary tool with respect to the current genetic and manual immunohistochemistry methods.

References

[1] Ali Z, Yousaf N, Larkin J (2013) Melanoma epidemiology, biology and prognosis. Eur J Cancer, Suppl 11:81–91. https://doi.org/10.1016/j.ejcsup.2013.07.012
[2] Hyams DM, Cook RW, Buzaid AC (2019) Identification of risk in cutaneous melanoma patients: Prognostic and predictive markers. J Surg Oncol 119:175–186. https://doi.org/10.1002/jso.25319