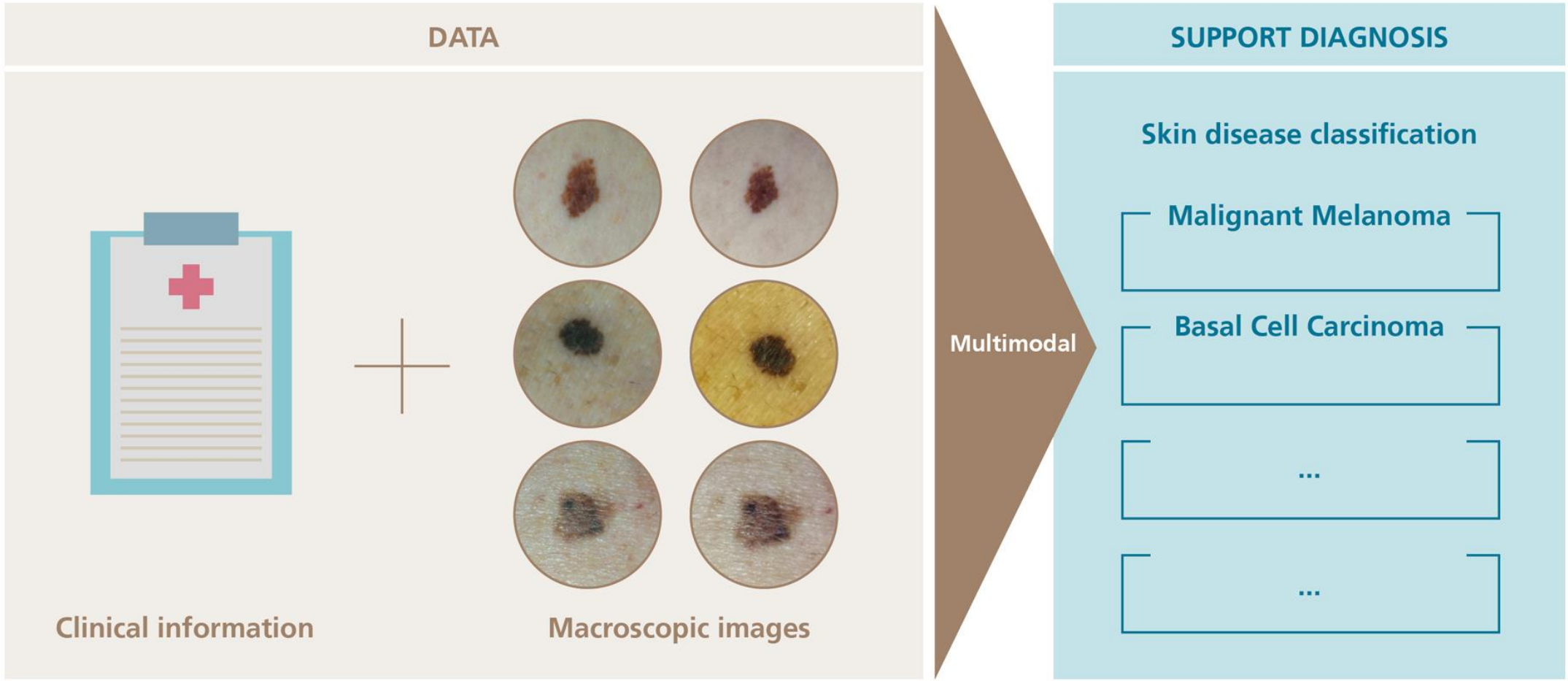
MultimodalDermaCAD - Classification of multimodal dermatological data



Nome: David Gonçalo Ferreira Oliveira

Orientador: Inês de Castro Dutra

Coorientador: Maria Vasconcelos and Tudor-Ionut Nedelcu.

Index

Motivation

- Skin cancer is one of the more dangerous conditions overseen in Dermatology, with several types existing:
  - Basal cell carcinoma (BCC), most common.
  - Melanoma (MEL), most dangerous.
  - Nevus (NEV), benign lesion
  - Seborrheic Keratosis (SK), benign lesion
- Diagnosing skin cancers is a complicated process.
- Screening has been used to preemptively identify skin cancer.
- Screening for skin cancer is neither recommended nor discouraged by various organizations such as the US preventive services task force.

Literature review

- Computer-aided diagnosis (CAD) systems for skin cancer have been under research since 1987.
- Deep learning is a subfield of artificial neural networks that has been on the rise since it manage to improve the results by a significant margin in the ImageNet competition challenge in 2012.
- A Convolutional Neural Network (CNN) is a type of Deep Neural Networks able to receive images as inputs and extract meaningful features from them.
- Recently, CAD system have been taking advantage of powerful CNN to achieve similar results to more classical methods.
- CAD systems utilizing CNNs have been able to surpass groups of dermatologists, especially new dermatologists, in diagnosis performance.
- Most studies do not take advantage of all the available data in the domain.
- Problems include the access to public datasets and the quantity of data for training, especially when DNN are used.

Objectives

- Improve detection of Skin cancer:
  - Investigate the impact of each modality to the quality of the skin cancer classification.
  - Investigate the impact of the fusion of the modalities in the quality of the skin cancer classification.
  - Investigate the impact of several techniques to the quality of the skin cancer classification.
  - Determine the viability of a simple architecture in skin cancer classification.
  - Investigate the impact of performing several combinations of modalities on the same model.

Datasets

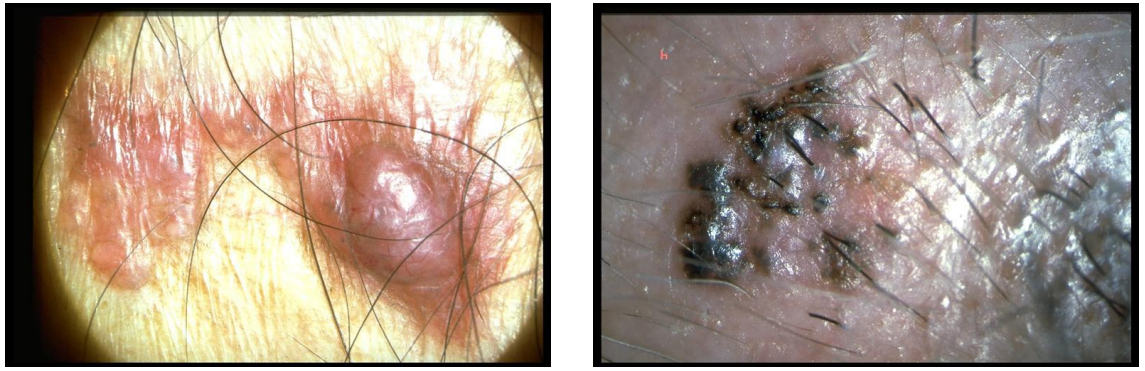
- Datasets with all data types is preferred:
  - Metadata:
    - Sex
    - Age
    - Location
    - …
  - Clinical images
  - Dermoscopic images.
- Of the 24 datasets, only the EDRA and the ISIC archive datasets have a combination of all the above data types.
- Work performed has been done so far on the EDRA dataset.

EDRA	Clinical Dermoscopic Metadata	1011	public
University of Tsukuba Hospital	Clinical	6009	private
SD-198	Clinical	6584	NA
Dermofit	Clinical	1300	public
HAM10000	Dermoscopic Metadata	10015	public
DermQuest	NA	NA	deactivated
ISIC Archive	Clinical Dermoscopic Metadata	NA	public

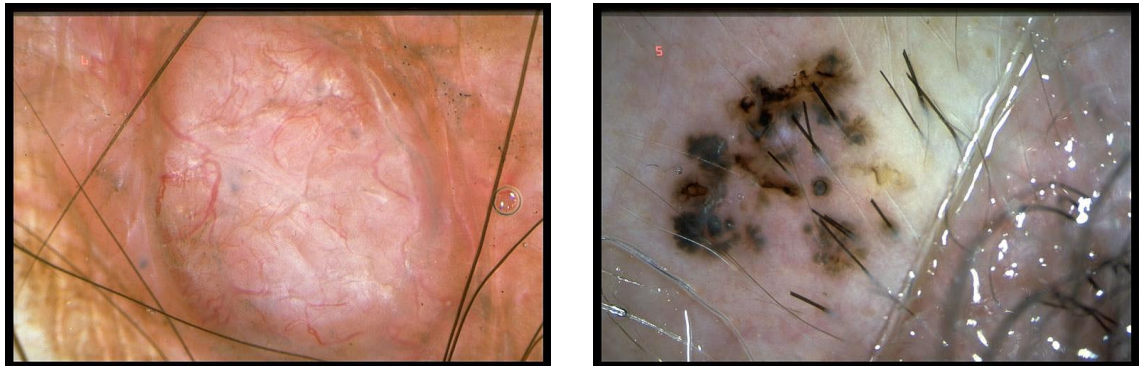
EDRA Dataset

- Descriprion:
  - dermoscopic images
  - clinical (macro) images
  - Metadata:
    - Elevation
    - Location
    - Sex
  - 7-point criteria annotations.
  - Skin lesion classification.
- 1011 total number of samples.
- 41/20/39% train/validation/test split.

Clinical



Dermoscopic



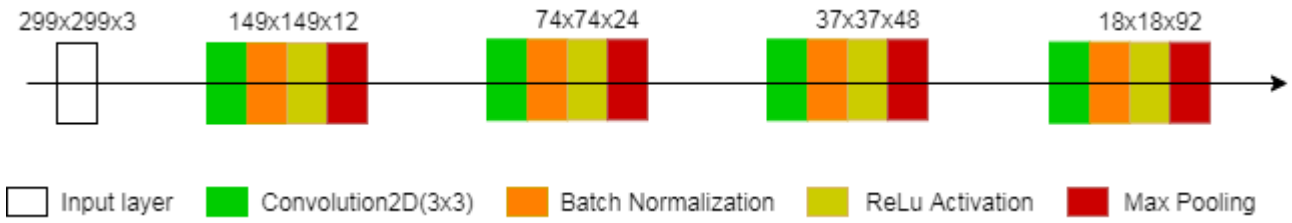
Methodology

- Using a small custom architecture for the CNN.
- Class balancing:
  - over-sampling for skin lesion classification
  - weighing for other learning tasks (when multitasking).
- Data augmentation.
- Train and Validation metrics:
  - Loss: Sparse Categorical Crossentropy $(-\sum_{i=1}^{\text{number of samples}} y\_\text{true}_i * \log y\_\text{pred}_i)$
  - Accuracy
- Test metrics:
  - Accuracy (ACC)
  - Sensitivity (SEN)
  - Specificity (SPC)
  - F1-score
  - Receiver Operating Characteristic (ROC) curve
  - Area under the curve (AUROC)

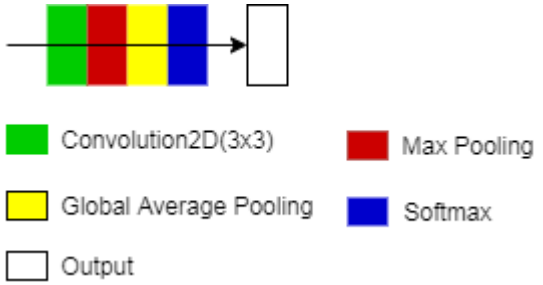
Methodology: Architecture

- The architecture is divided into several modules, with each experiment utilizing the modules as needed.
- The modules are the same for every experiment.

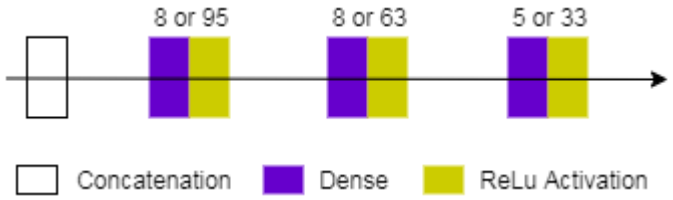
- Feature extraction module



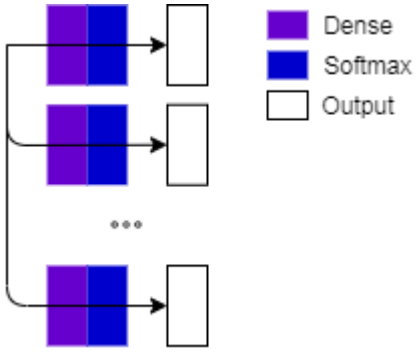
- Image classification module



- fusion module



- multitask classification module



Methodology: Experiments

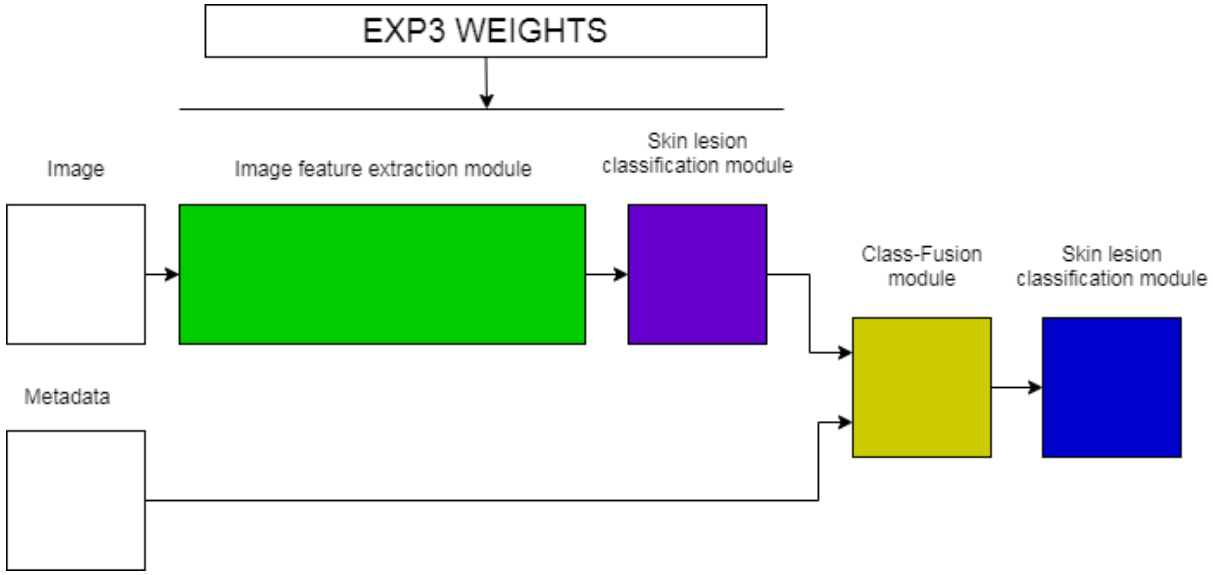
- A Series of experiments are performed from a simple solution to a more complex one, involving all the modalities:
  - Exp1: Img
  - Exp2: ImgMd_CF
  - Exp3: Img_MT7pts
  - Exp4: ImgMd_CF_TransfL7pts
  - Exp5: Img_MtMd
  - Exp6: ImgMd_CF_TransfLMd
  - Exp7: ImgMd_FF
  - Exp8: ImgMd_FF_MT7pts
  - Exp9: 2Img_FF_MT7pts
  - Exp10: 2ImgMd_FF_MT7pts
  - Exp11: 2ImgMd_CombFF_MT7pts
- Due to the extension of the results, only a portion is presented next.

Results

Averages	Clinical					Dermoscopic				
	ACC	SEN	SPC	F1	AUC	ACC	SEN	SPC	F1	AUC
EXP1	0.44	0.374	0.846	0.336	0.684	0.59	0.464	0.882	0.434	0.806
EXP2	0.46	0.372	0.848	0.344	0.690	0.45	0.466	0.854	0.396	0.772
EXP3	0.45	0.316	0.844	0.300	0.654	0.60	0.476	0.878	0.436	0.796
EXP4	0.54	0.414	0.866	0.400	0.698	0.64	0.474	0.882	0.466	0.800
EXP5	0.47	0.364	0.850	0.342	0.682	0.57	0.456	0.872	0.418	0.798
EXP6	0.54	0.400	0.866	0.388	0.696	0.59	0.496	0.880	0.446	0.786
EXP7	0.52	0.394	0.860	0.382	0.742	0.60	0.484	0.886	0.444	0.814
EXP8	0.56	0.450	0.872	0.438	0.762	0.61	0.500	0.882	0.470	0.790
EXP9	0.63	0.456	0.882	0.450	0.762					
EXP10	0.65	0.520	0.886	0.508	0.772					
EXP11(I)	0.44	0.342	0.842	0.316	0.658	0.58	0.448	0.878	0.408	0.756
EXP11(I+M)	0.52	0.408	0.864	0.384	0.732	0.62	0.534	0.888	0.488	0.804
EXP11(ALL)	0.65	0.456	0.890	0.450	0.758					

Results: dermoscopic experiment 4

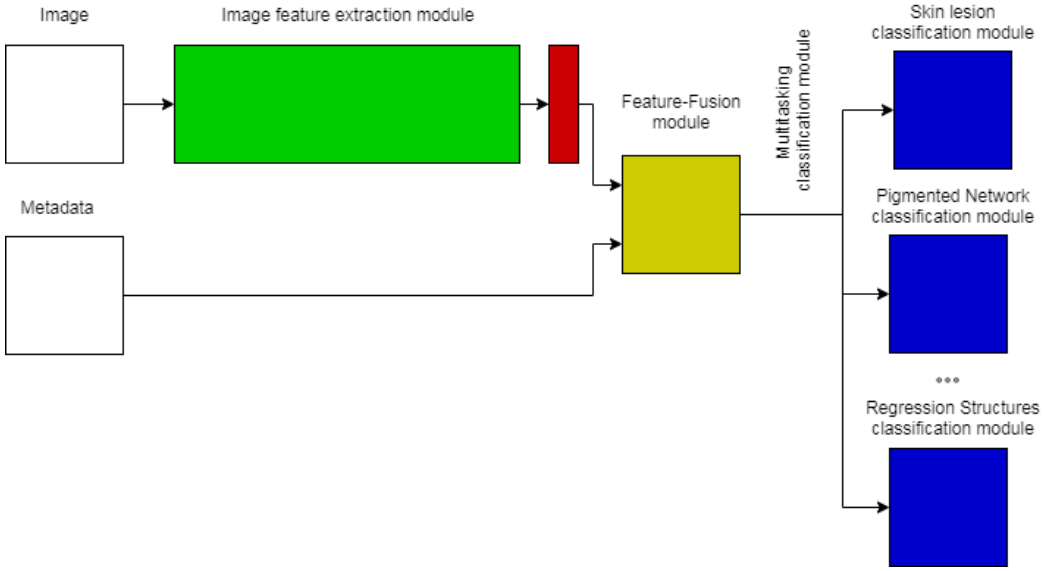
- Transfer learning the knowledge from the third experiment.
- Class-fusion.
- Large improvement to the results compared to the dermoscopic class-fusion (0.45 ACC from exp2, 0.59 ACC from exp1 and 0.60 from exp3).



Results	DImgMd_CF_TransfL7pts				
	ACC	SEN	SPC	F1	AUC
BCC	0.64	0.31	0.98	0.36	0.78
NEV		0.81	0.65	0.78	0.81
MEL		0.41	0.92	0.50	0.70
MISC		0.68	0.89	0.51	0.89
SK		0.16	0.97	0.18	0.82
AVG		0.474	0.882	0.466	0.8

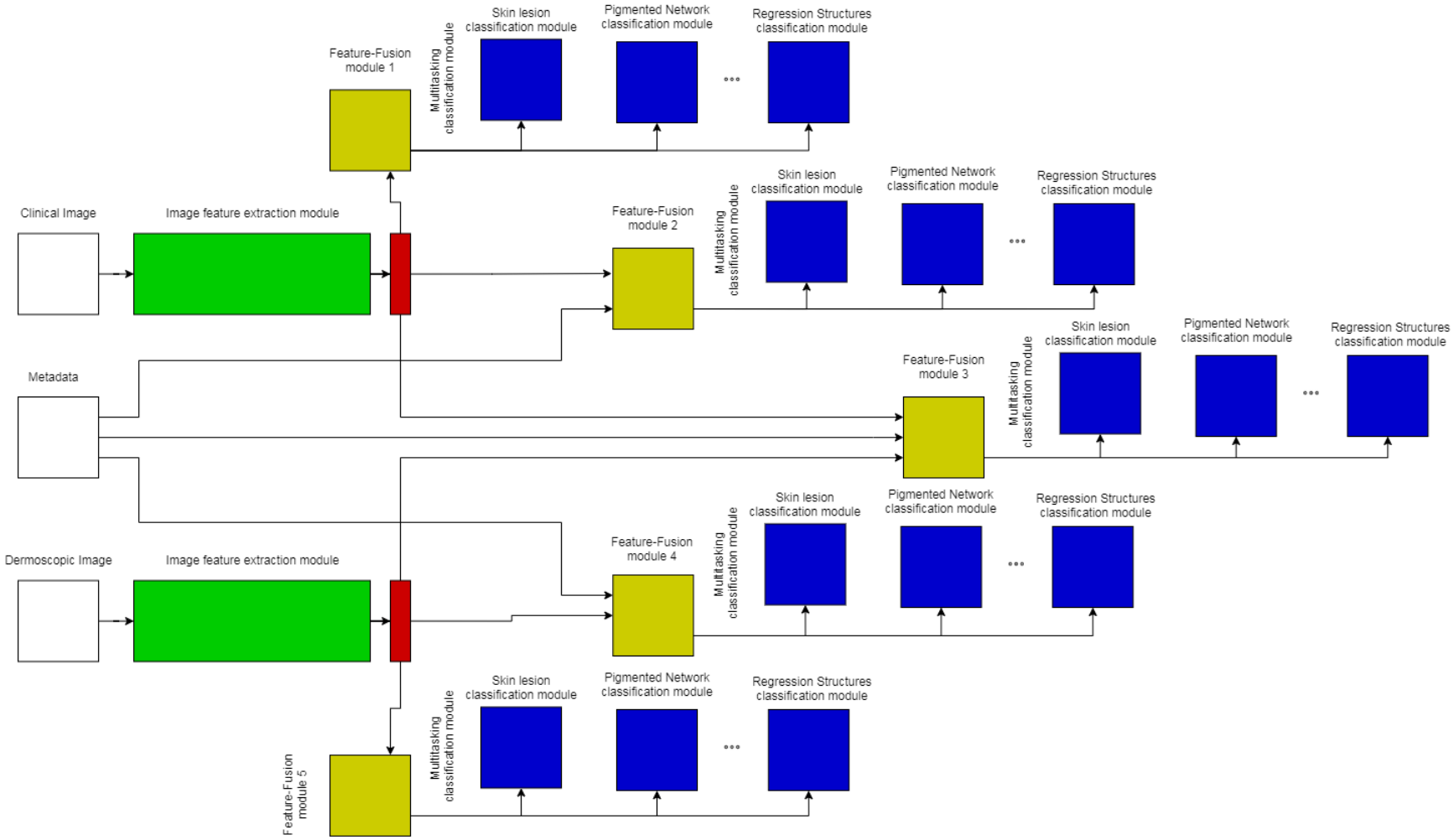
Results: dermoscopic experiment 8

- Feature-fusion.
- Multitasking.
- Improvement to the results compared to the dermoscopic class-fusion (0.45 ACC from exp2 and 0.60 ACC from exp7).



Results	DImgMd_FF_MT7pts				
	ACC	SEN	SPC	F1	AUC
BCC	0.61	0.38	0.98	0.39	0.81
NEV		0.76	0.68	0.75	0.8
MEL		0.35	0.93	0.45	0.75
MISC		0.75	0.85	0.48	0.86
SK		0.26	0.97	0.28	0.73
AVG		0.5	0.882	0.47	0.79

Results: dermoscopic experiment 11 (architecture)



Results: dermoscopic experiment 11

- Various combinations of modalities.
- Feature-fusion.
- Multitasking.
- Trained faster (compared to the summed total from training each combination individually).
- Generally obtains slightly lower results (0.65 ACC from exp10).
- The dermoscopic image and metadata combination obtained better results than its equivelent (0.61 ACC from exp8 and 0.62 ACC from exp11(DM)).

Results	dermoscopic models				
	ACC	SEN	SPC	F1	AUC
EXP3	0.60	0.476	0.878	0.436	0.796
EXP8	0.61	0.5	0.882	0.47	0.79
EXP10 (ALL)	0.65	0.52	0.886	0.508	0.772
EXP11(I)	0.58	0.448	0.878	0.408	0.756
EXP11 (I+M)	0.62	0.534	0.888	0.488	0.804
EXP11 (ALL)	0.65	0.456	0.89	0.45	0.758
EXP11 (Kawahara)		0.604	0.91		0.896

Conclusions

- Dermoscopic images give more information to the model.
- Metadata can be difficult to fuse with dermoscopic images.
- Feature-fusion obtains better results than class-fusion.
- Metadata (location and elevation of the lesion and sex of the patient) has a high bias for the BCC cancer.
- Multitasking introduces a smaller influence on the model from the modality being multitasked.
- Transfer learning can also be used to gradually introduce modalities during the training of the model, producing better results from the fusion of the modalities.
- A model focused on a single combination of modalities generally obtains the best results, as opposed to a model with various combinations of modalities and a large number of learning tasks.
- The combination model it a good approach to investigate various possibilities simultaneously, as it trained faster while obtaining a slightly lower performance.
- A simple network can investigate ideas in a timely manner, but does not produce the best results.

Future work

- Investigating the gradual introduction of more modalities.
- Investigating a more direct approach to the gradual introduction of more modalities.
- Investigating if the order with which the modalities are introduced is relevant.
- Investigating the reasons for the model obtaining better results by being introduced to a modality first. This however is more fitted for the "Explainable AI" research field.

Questions?