



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Biomedical Signal Processing and Control

journal homepage: [www.elsevier.com/locate/bspc](https://www.elsevier.com/locate/bspc)



### A Multi-group similarity-decoding-based method for deep model ensembling applied in the microcalcification classification on digital mammograms

Liuwen Zou<sup>a,1</sup>, Jiarui Sun<sup>a,1</sup>, Jingyu Li<sup>a</sup>, Hui Tang<sup>a,\*</sup>, Guangquan Zhou<sup>b</sup>, Gouenou Coatrieux<sup>c</sup>, Jean-Louis Coatrieux<sup>d</sup>, Yang Chen<sup>a,e,f</sup>, Yanni Jiang<sup>g,\*</sup>

<sup>a</sup> The Laboratory of Image Science and Technology, Southeast University, Nanjing, Jiangsu, China

<sup>b</sup> State Key Laboratory of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing, Jiangsu, China

<sup>c</sup> The IMT Atlantique, Inserm, LaTIM UMR1101, Brest 29000, France

<sup>d</sup> The National Institute for Health and Medical Research, F-35000 Rennes, France

<sup>e</sup> Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, Southeast University, Nanjing, Jiangsu, China

<sup>f</sup> The Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Ministry of Education, Nanjing 210096, China

<sup>g</sup> The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, China

答 辯 人: 邹刘文

指 导 老 师: Jean-Louis Coatrieux,  
唐慧, 周光泉, 陈阳

2024/04/16



# 目录

01

## 研究背景

- 1.1 研究问题及意义
- 1.2 研究难点

## 研究现状

- 2.1 基于特征工程的传统方法
- 2.2 基于深度学习的方法
- 2.3 基于深度集成学习的方法

02

03

## 论文提出的模型

- 3.1 提出的MG-SDM及其在分类模型中的使用
- 3.2 消融实验及其结果
- 3.3 对比实验及其结果
- 3.4 扩展实验及其结果

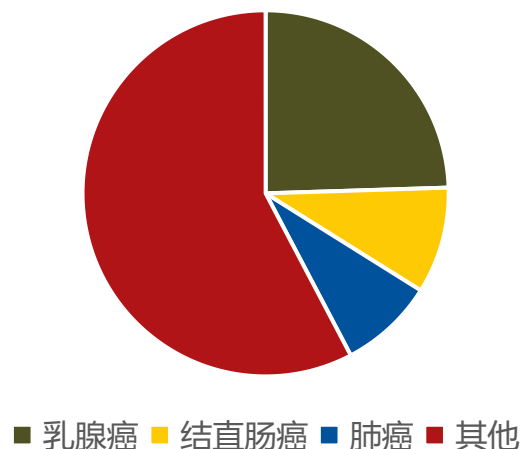
## 研究复盘

- 4.1 研究过程中的启发

04

# 1.1 研究问题及意义

《World Cancer Report 2020》中女性各种癌症的案例比例



A) 发病率和致死率均最高

B) 数字钼靶成像 (DM, Digital Mammography) 因成本低和高敏感度 而成为乳腺癌早期筛查的主要手段

乳腺影像报告和数据系统 BI-RADS 是由美国放射学院提出的DM临床判断标准

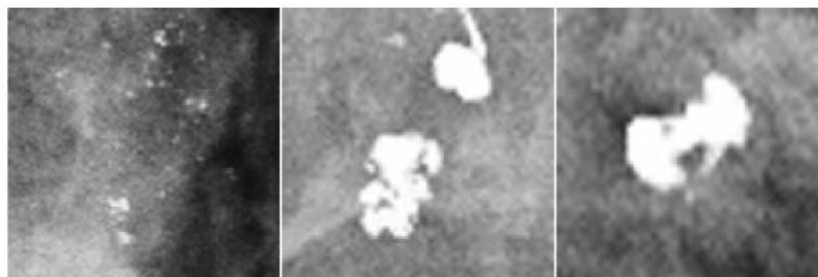
表 1- 1 BI-RADS 报告系统描述

BI-RADS 分类		影像诊断描述
0		根据现有的检查片无法准确判断分级, 需结合病史或加做其他影像检查才能判断
1		正常乳腺或者存在正常改变的乳腺
2		能够明确判断为乳腺良性病变
3		可能为良性改变(恶性可能性小于等于 2%, 建议短期随访复查)
4	4a	轻度疑似恶性, 恶性可能性大于 2% 小于或等于 10%, 一般需要外科处理(如活检等)。常见的有乳腺纤维腺瘤, 有一定恶变可能的乳腺囊肿及乳腺脓肿等。
	4b	中度疑似恶性, 恶性可能性大于 10% 小于或等于 50%, 但也可能是乳腺纤维腺瘤, 脂肪坏死或者导管内乳头状瘤, 需要病理检查等外科处理。
	4c	疑似恶性, 恶性可能性大于 50% 小于或等于 95%, 通常表现为乳腺高密度肿物边界欠清晰, 需要病理活检确认。
5		高度疑似恶性, 恶性可能性大于 95%, 通常表现为乳腺高密度肿物边界欠清晰, 或出现大片群集微小钙化灶及乳腺肿物呈明显毛刺征, 需要尽快安排外科手术。

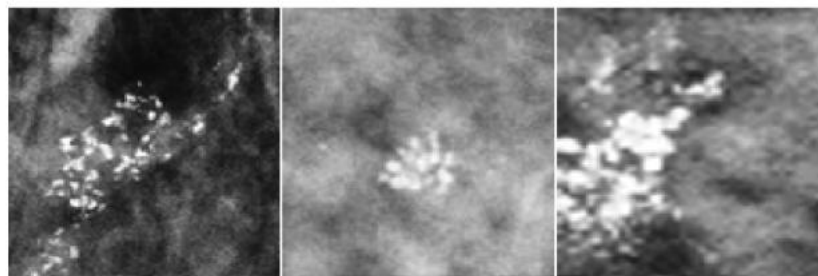
“对BI-RADS评级为4的DM图片进行良恶性分类” 问题的研究, 对于提高诊断率和减少病人不必要的活检痛苦具有积极意义

## 1.2 研究的难点

乳腺钼靶图像上良恶性分类任务的难点是：



(a) Benign Samples.



(b) Malignant Samples.

尺寸更大

分布更分散

形状更粗糙（圆形状、边状、爆米花状）

尺寸更小

分布更密集

形状更精细（蠕虫状、铸造状、树枝状）

Fig. 1. Benign and malignant samples for mammography.

钼靶图像中的**白色点状或斑块状阴影**  
被称作**钙化点 (MCs, Microcalcifications)** ,  
其是乳腺钙化这种病理性变化在乳腺内  
沉积的小点或者斑块

钙化点 (MCs)  
在shape和distribution  
上极具多样性

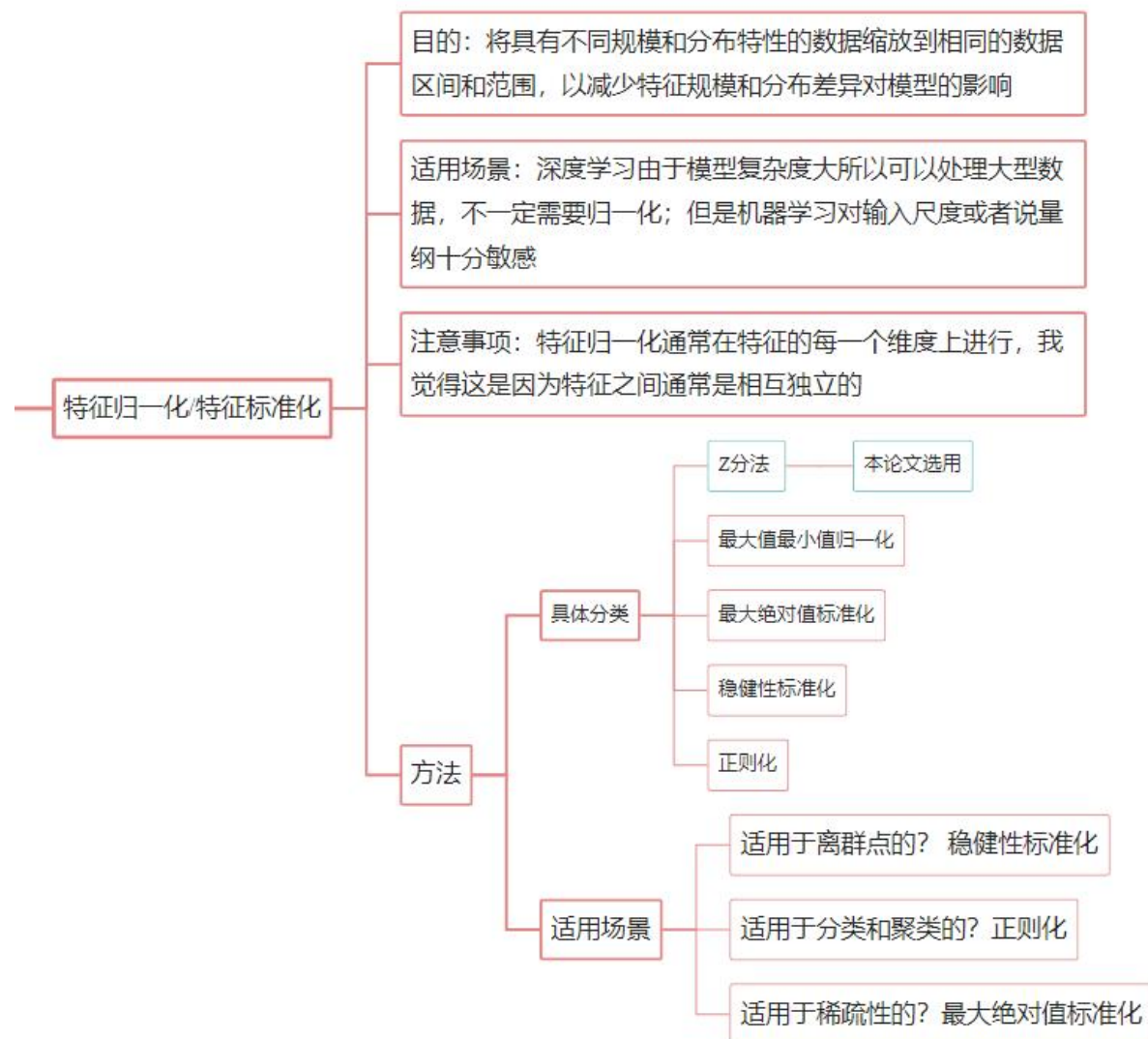
模型难以实现较高的  
泛化性或者鲁棒性

## 2.1 基于特征工程的传统方法





## 2.1 基于特征工程的传统方法



## 2.1 基于特征工程的传统方法

### 特征降维与分析

#### 特征降维/选择

##### 目的

- 防止高维数据带来的过拟合
- 排除冗余特征，提高模型精度
- 简化模型，提高模型的可解释性和可读性

### 特征过滤(1):

#### 单因素方差分析法 (F检验或者变异数分析)

##### 目的

通过比较多个组别的数据均值是否一致来判断这些组别之间是否有统计学的显著性差异

##### 方法

- 如果各组数据近似服从正态分布且方差齐次，偏态不严重，则使用参数型检验方法ANOVA
- 如果偏态严重，则使用非参数检验方法Kruskal-Wallis

##### 本论文中的做法

- 由于之前使用了Z分法使得特征服从标准正太分度，偏态或者说偏度应该为0，故直接使用ANOVA
- 将所有数据分为良性和恶性两组，然后计算关于每个特征的这两个组别的F统计值

F统计值的计算 = 观测量的组间均方 / 观测量的组内均方

组间均方 = 组间平方和 / 组间自由度

组内均方 = 组内平方和 / 组内自由度

### 特征过滤(2):

#### 皮尔森相关系数法

这里就是针对每一个特征，删除与其线性相关性最强的特征

### 特征过滤(3):

#### 基于普通最小二乘法 (OLS) 的后向特征筛选

使用多元线性回归方程来拟合预测标签Y，通过最小二乘法来求得最佳拟合的多元线性回归方程的参数值

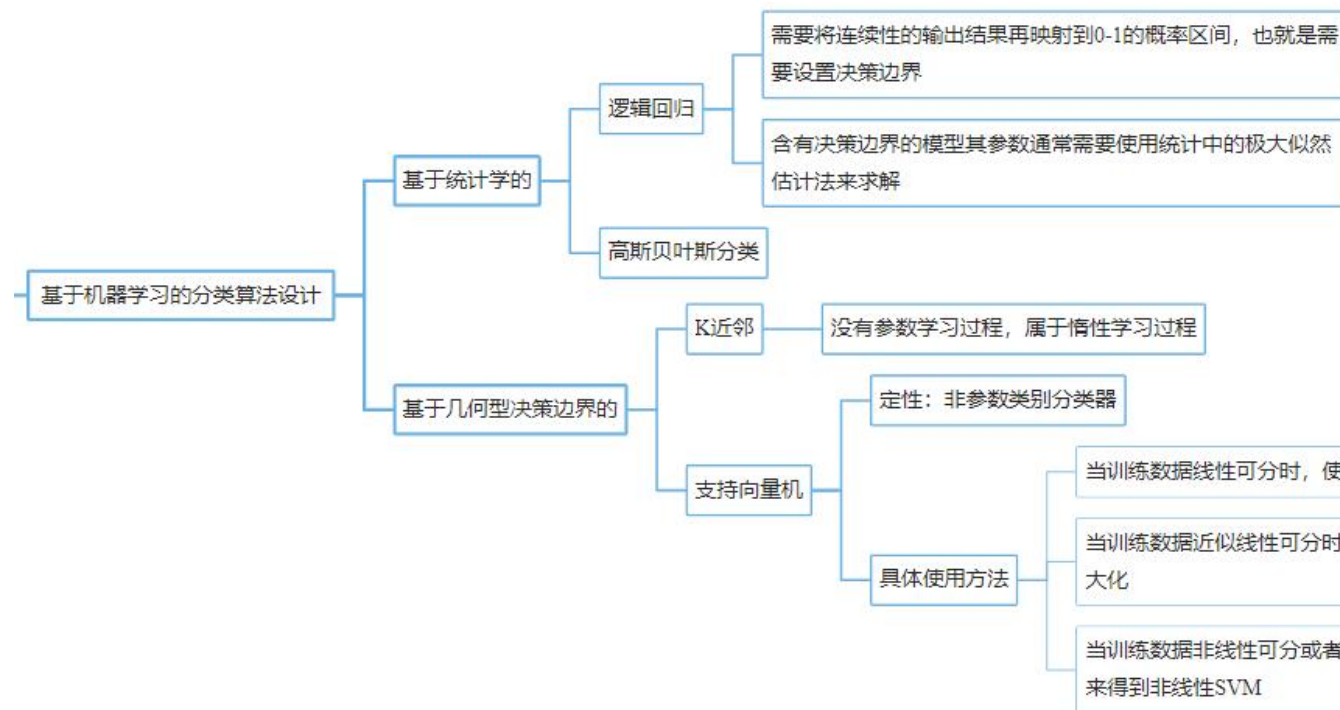
备注：后向特征筛选的含义就是先使用OLS确定每个特征变量对应的最佳权重值，然后再对输出结果做分析来筛选掉一个特征，然后再对筛选后的特征组合再次重复上述过程，如此反复。后向的意思，是从结果来反向筛掉输入

### 特征过滤(4):

#### 基于XGBoost的递归特征筛选

递归特征筛选算法：使用一种基于机器学习的分类模型来对问题进行分类，可以利用该分类模型获取不同特征的关于重要性度量的一个值，然后利用这些值来筛掉特征；不断递归进行下去，直到筛选出目标数量的特征集

## 2.1 基于特征工程的传统方法



Details of ten selected features.

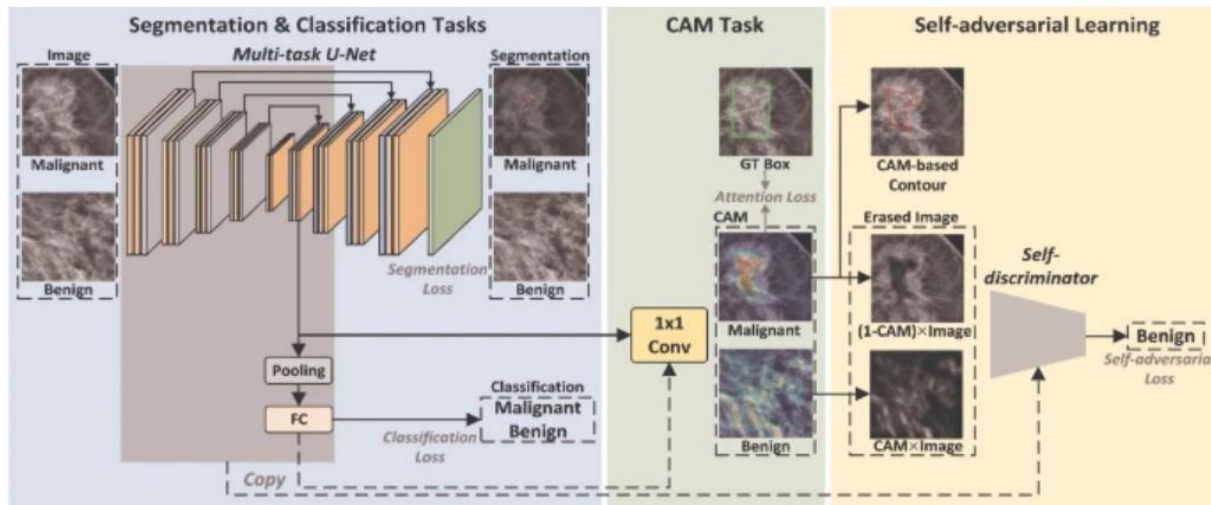
Feature No	Feature Name
1	original-glszm-LowGrayLevelZoneEmphasis
2	log-sigma-1-mm-3D-glszm-ZoneVariance
3	log-sigma-2-mm-3D-glszm-ClusterProminence
4	log-sigma-3-mm-3D-glszm-SumSquares
5	log-sigma-3-mm-3D-glszm-SmallAreaLowGrayLevelEmphasis
6	WaveletLH-glszm-GrayLevelNonUniformityNormalized
7	wavelet-LH-glszm-SmallAreaEmphasis
8	wavelet-LL-firstorder-90Percentile
9	wavelet-LL-glszm-GrayLevelNonUniformityNormalized
10	MaximumVertexDegree-27



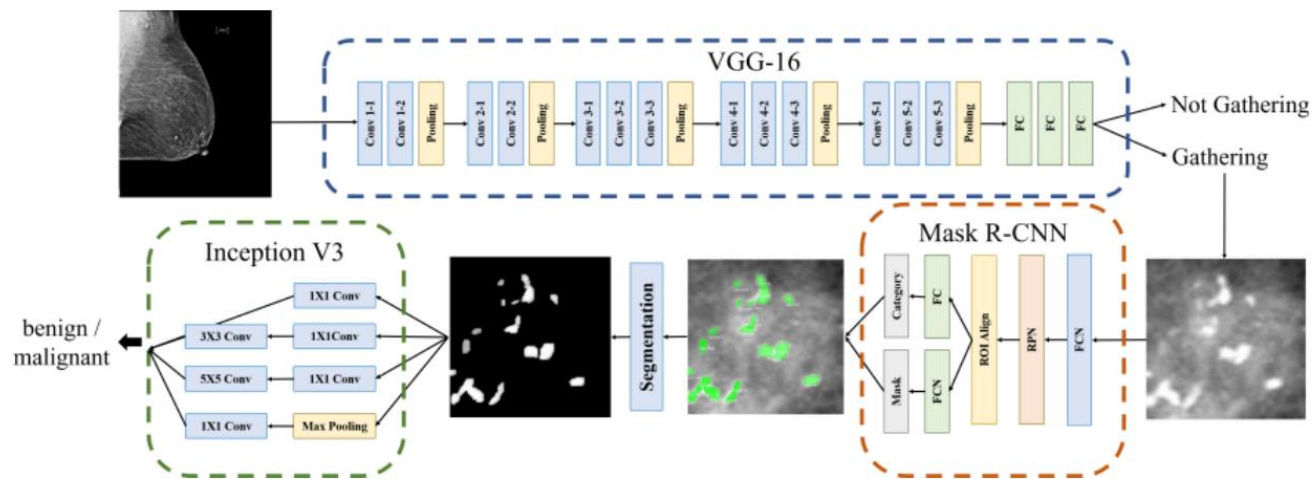
该类方法的不足：**过程繁琐且耗时长**、模型性能容易受到数据集本身的影响



## 2.2 基于深度学习的方法



将钙化点分割、钙化点良恶性分类和钙化点检测的三个任务同时在一个网络中完成



首先vgg16 and mask rcnn来定位钙化点MCs, 然后利用inception\_v3来进行良恶性分类

该类方法的不足: 容易过拟合, 且因为复杂模型的参数量较大而导致泛化性能一般

## 2.3 基于深度集成学习的方法



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

### Engineering Applications of Artificial Intelligence

journal homepage: [www.elsevier.com/locate/engappai](http://www.elsevier.com/locate/engappai)



Survey paper

#### Ensemble deep learning: A review

M.A. Ganaie<sup>a</sup>, Minghui Hu<sup>b</sup>, A.K. Malik<sup>a</sup>, M. Tanveer<sup>a,\*</sup>, P.N. Suganthan<sup>b,c,\*</sup>

<sup>a</sup> Department of Mathematics, Indian Institute of Technology Indore, Simrol, Indore, 453552, India

<sup>b</sup> School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore

<sup>c</sup> KINDI Center for Computing Research College of Engineering, Qatar University, Qatar



#### ARTICLE INFO

##### Keywords:

Ensemble learning  
Deep learning

#### ABSTRACT

Ensemble learning combines several individual models to obtain better generalization performance. Currently, deep learning architectures are showing better performance compared to the shallow or traditional models. Deep ensemble learning models combine the advantages of both the deep learning models as well as the ensemble learning such that the final model has better generalization performance. This paper reviews the state-of-art deep ensemble models and hence serves as an extensive summary for the researchers. The ensemble models are broadly categorized into bagging, boosting, stacking, negative correlation based deep ensemble models, explicit/implicit ensembles, homogeneous/heterogeneous ensemble, decision fusion strategies based deep ensemble models. Applications of deep ensemble models in different domains are also briefly discussed. Finally, we conclude this paper with some potential future research directions.

**集成学习能够提升泛化性的原因：**通过对多个模型的集成，方差和偏差都能有所降低

**decision fusion中naive averaging的不足：**每个模型都有一个对应的权重，如果存在biased learner赋上较大权重，则对总体结果影响大

**data-adaptive的权重计算：**第一类是数据驱动的，如在贝叶斯最优分类器中，每个基础模型都被表示为一个假设，并且通过最大化期望的对数后验概率来获得权重；第二类是模型驱动的，如叠加泛化利用交叉验证数据和最小二乘法在非负约束下帮助元模型学习权重。然而，这些实现方式依赖于复杂的数学或训练技巧。

# 3.1 提出的MG-SDM及其在分类模型中的使用

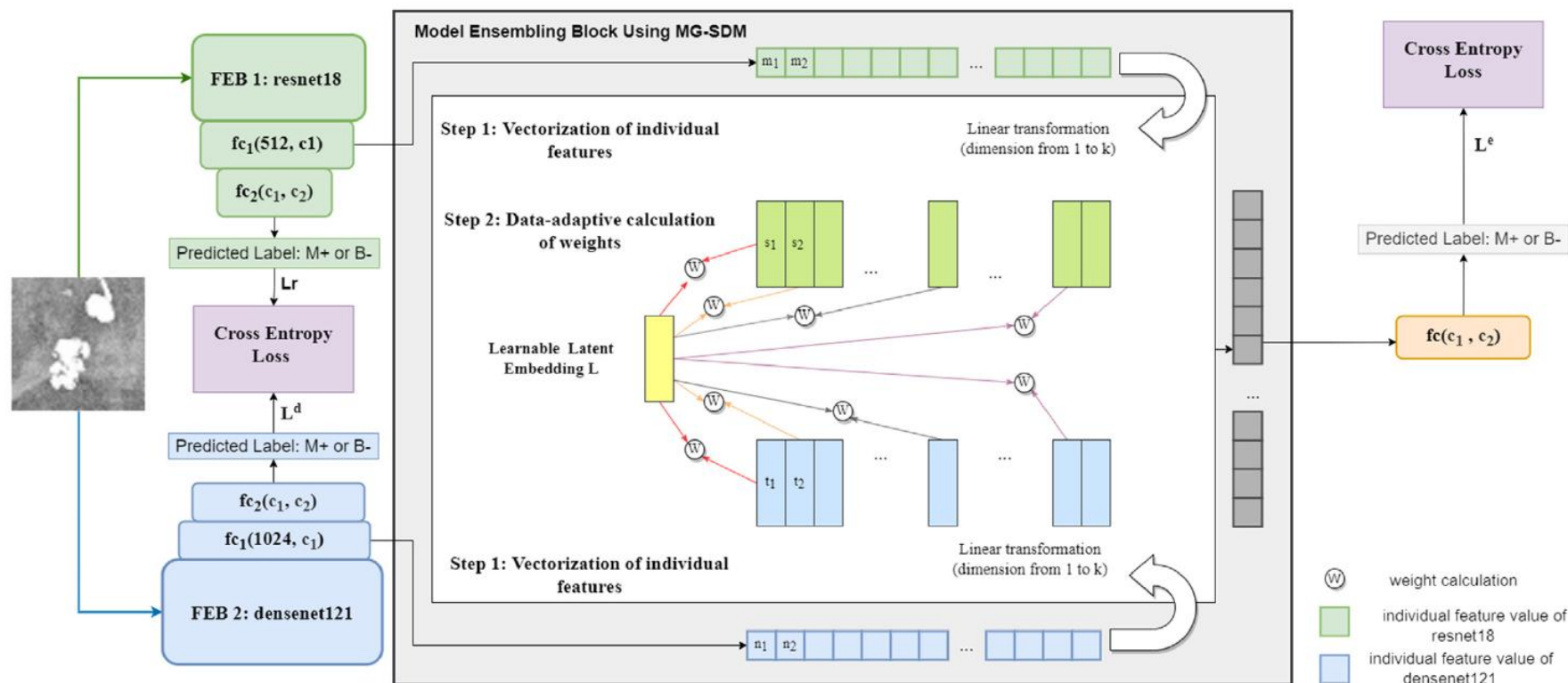


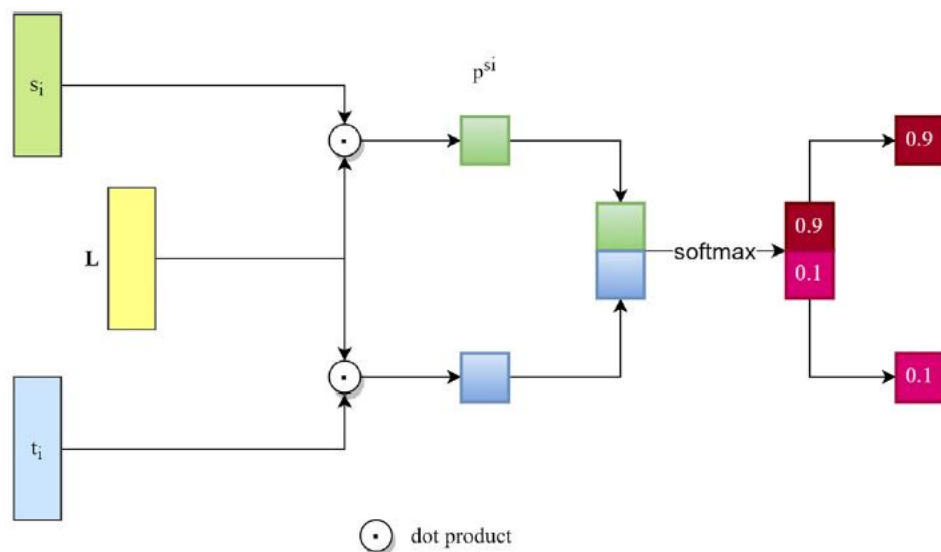
Fig. 2. A high-level overview of the proposed model. FEB is the abbreviation of feature extraction block.  $c_1$  is the intermediate dimension,  $c_2$  is the total number of categories ( $c_2 = 2$  in our case). Mathematically we have  $1024 > 512 > c_1 > c_2$ . The superscript  $r$  refers to the resnet18 branch,  $d$  refers to the densenet121 branch and  $e$  refers to the model ensembling branch.  $m$  is the feature vector extracted by resnet18;  $n$  is the feature vector extracted by densenet121; and  $f$  is the fused vector from  $m$  and  $n$ . In “Predicted Label: M+ or B-”, “M” stands for “Malignant” while “B” stands for “Benign”, and “+” means positive category while “-” means negative category.

将resnet18和densenet121  
都从1个FC层改成2个FC层

(Model Fusion to Feature-level Fusion) resnet18和densenet121各自在**第一层FC后**输出的两个等长特征向量v1和v2，通过使用MG-SDM机制的模型融合块进行融合，融合后的特征向量与v1等长；融合后的向量进行最终的FC和分类预测输出

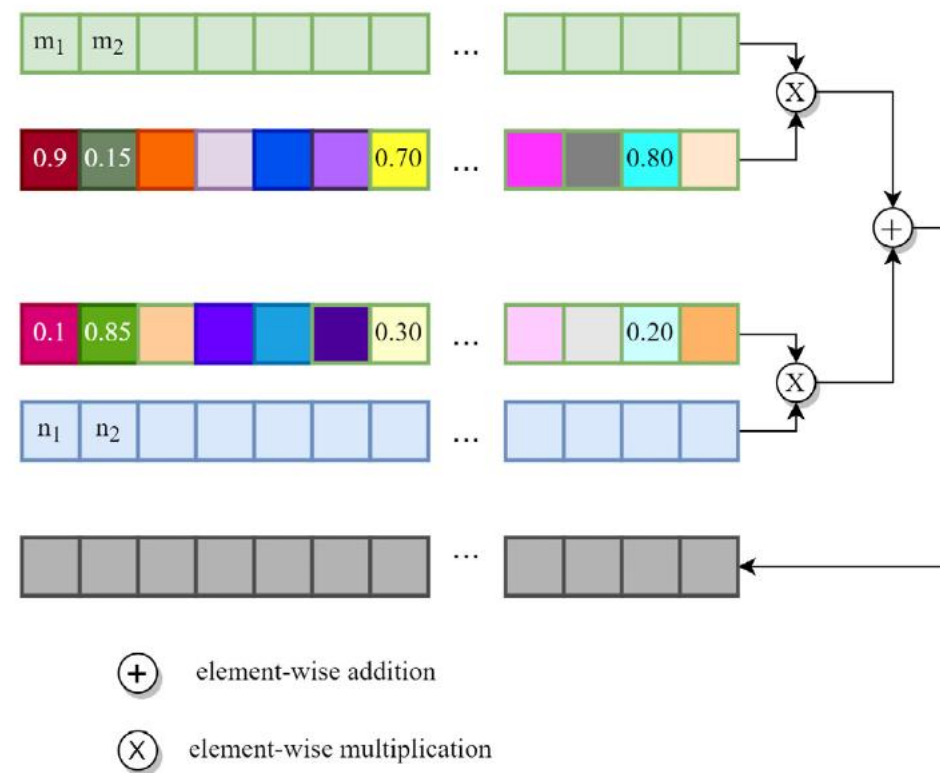
(Multi-output) resnet18和densenet121各自在**第二层FC后**同样输出分类结果，用作两个深监督分支

# 3.1 提出的MG-SDM及其在分类模型中的使用



L是可学习的潜在向量，其语义是**most task-related**的特征；

在**每组内利用SDM**（Similarity-decoding-based Mechanism）来计算组内两个不同特征的权重



这里**组别总数 (Total Group Number)** 等于特征向量的长度，待特征向量上所有位置处对应的组都计算得到归一化的权重后，按位置进行特征值的加权求和，即可得到融合后的向量



## 3.2 消融实验及其结果

**数据集来源：**南京医科大学附属第一医院

**数据集构成：**240个病人的数据，病人可能只有中侧斜位视图或者头尾位视图或者两者均有，所以一共有428张图片；其中，218张被标记为恶性，210张被标记为良性

**数据集处理：**没有进行数据增强，因为裁剪旋转都可能会丢失部分钙化点，而其他增强方式经过实验验证无法提升模型效果；**为了避免数据泄露**，同一个病人的所有视图均放置在训练/验证/测试中的同一个子集里；按照8:1:1固定划分数据集

**消融实验结果：**

Model	Test				
	AUC	Acc. (%)	Recall (%)	Prec. (%)	F1
proposed	0.8853	83.72	77.27	89.47	0.8293
resnet18	0.7424	72.09	72.73	72.73	0.7273
densenet121	0.8528	72.09	54.55	85.71	0.6667
proposed w/o multi-output	0.8030	72.09	63.64	77.78	0.7000

**评价指标：**F1是综合性指标，需要重点关注

$$F1 = \frac{2 * Recall * Precision}{Recall + Precision}$$

## 3.3 对比实验及其结果

Comparison type	Test				
	AUC	Acc. (%)	Recall (%)	Prec (%)	F1
resnet18 and densenet121 w concatenation fusion	0.8485	79.07	68.18	88.24	0.7692
svm	0.7403	53.49	95.45	52.50	0.6774
random forest	0.7662	65.12	63.64	66.67	0.6512
knn	0.7089	62.79	77.27	60.71	0.6800
FDFFE	0.6407	58.14	68.18	57.69	0.6250
vgg11	0.8355	69.77	86.36	65.52	0.7451
vgg16	0.7749	69.77	59.09	76.47	0.6667
efficientnet-b0	0.7597	72.09	63.64	77.78	0.7000
alexnet	0.7035	58.14	63.64	58.33	0.6087

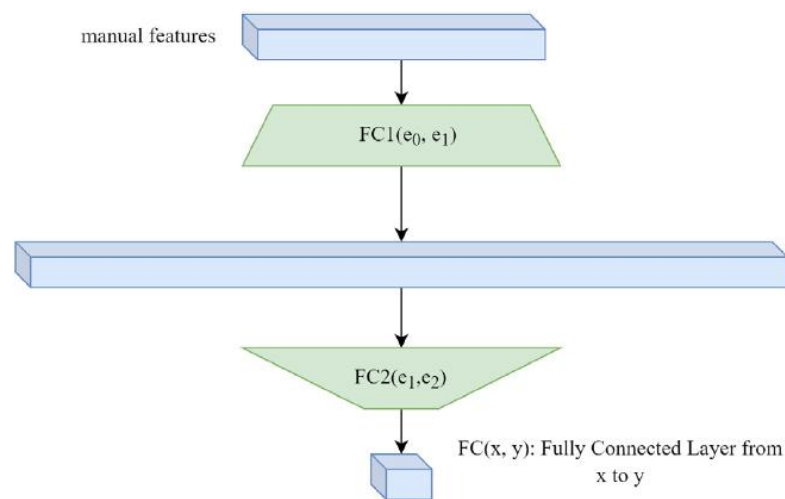


Fig. 10. The structure of the FDFFE model.  $e_0$ ,  $e_1$ , and  $e_2$  are all dimension values, with  $e_0 = 10$  and  $e_2 = 2$  being fixed.

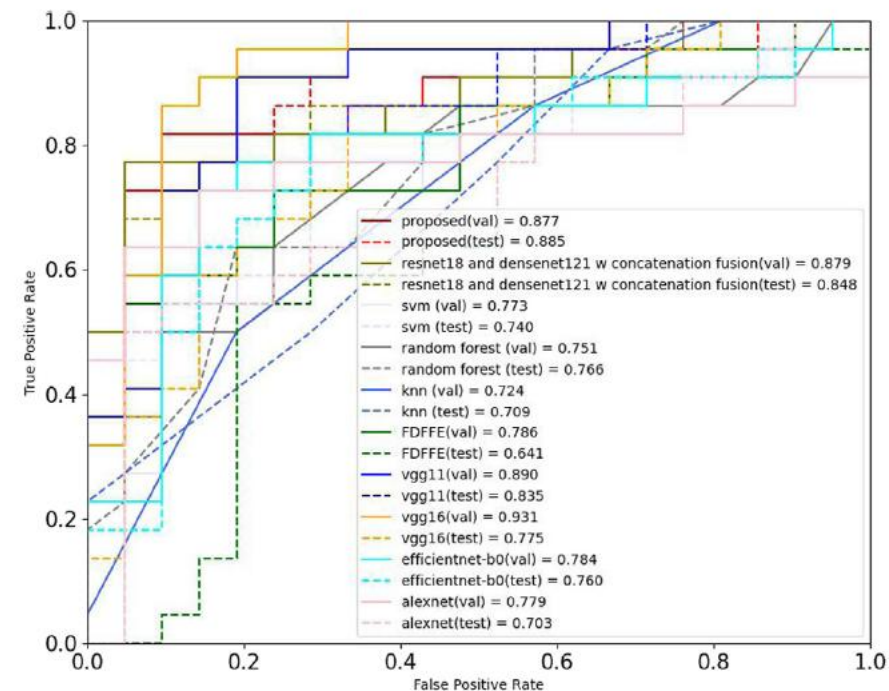
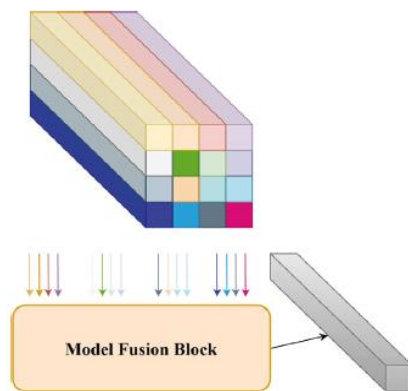
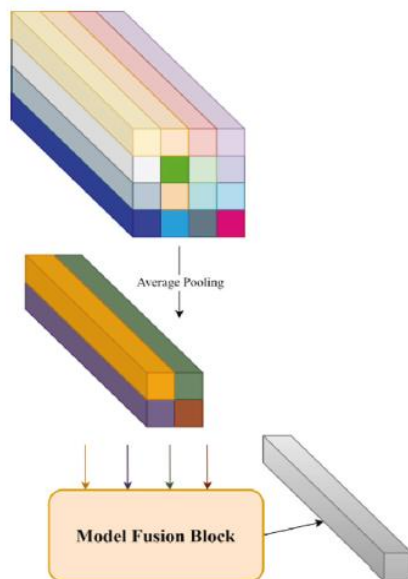


Fig. 8. The receiver operating characteristic curve (ROC) in quantitative comparison.

### 3.4 拓展实验及其结果



(a) resnet18 with weighted pooling.



(b) densenet121 with weighted pooling.

Fig. 11. An illustration of how resnet18 and densenet121 use weighted pooling respectively.

pooling + MG-SDM => weighed\_pooling

resnet18: 直接取代max\_pooling

densenet121: 先下采样为1/2, 再取代max\_pooling; 这是因为densenet特征图的维度过高, 会导致MG-SDM的参数量也相应变高

### 3.4 拓展实验及其结果

Model	Test				
	AUC	Acc. (%)	Recall (%)	Prec.(%)	F1
resnet18 w weighted pooling	0.9091	83.72	81.82	85.71	0.8372
densenet121 w weighted pooling	0.8550	83.72	81.82	85.71	0.8372

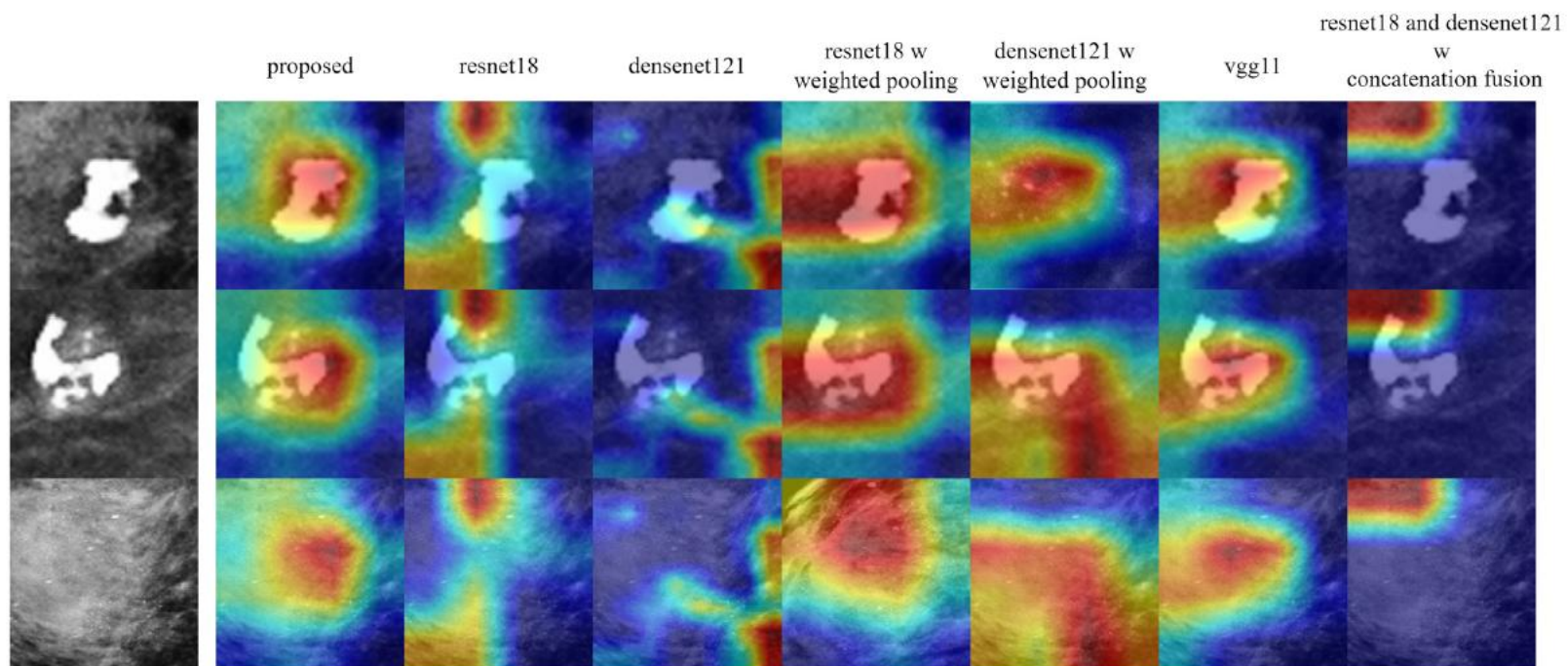


Fig. 12. The grad-cam visualizations for 7 models on three benign samples. If an image area has a darker red, the model will give it higher attention.



# 3.4 拓展实验及其结果

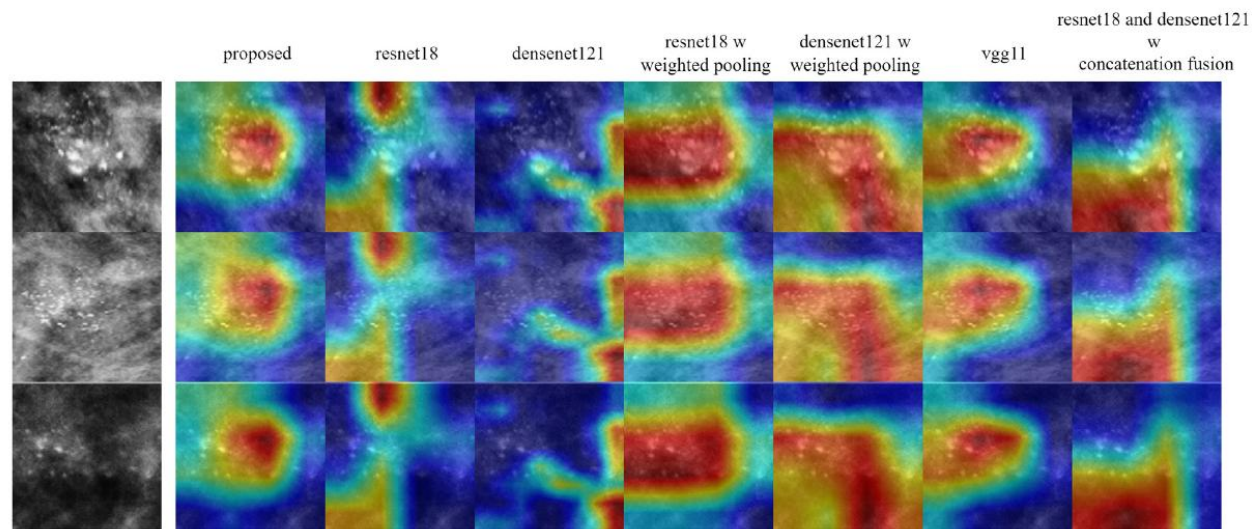
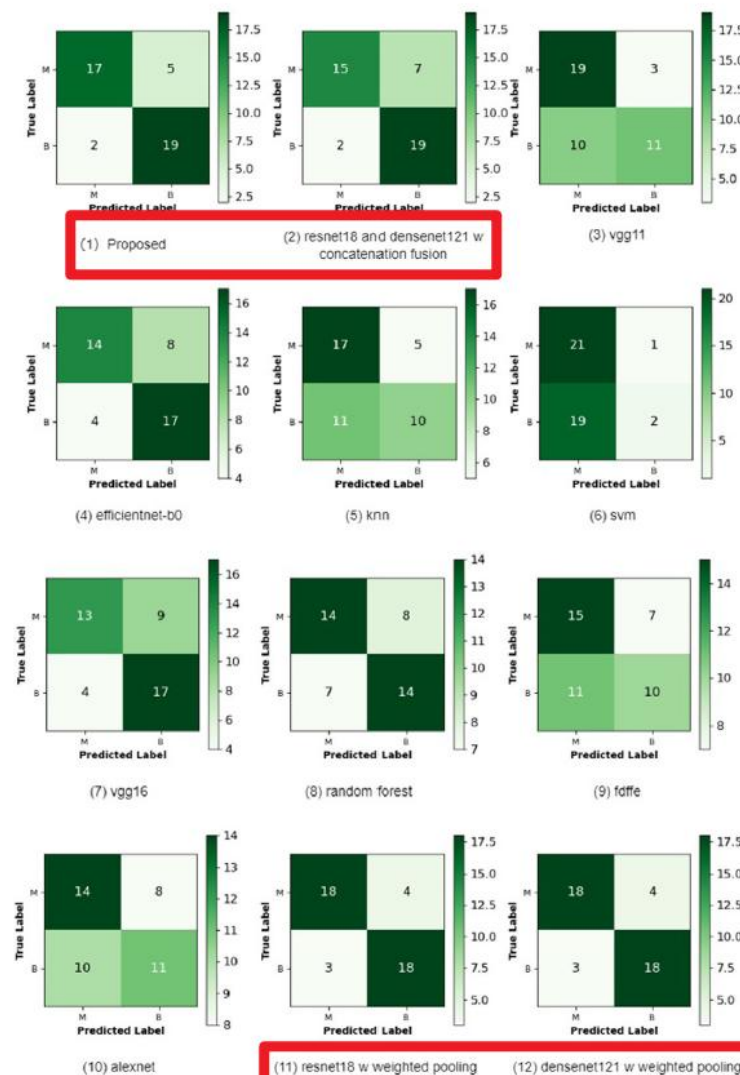


Fig. 13. The grad-cam visualizations for 7 models on three malignant samples.



## 4.1 研究过程中的启发

在具体的应用领域，并非越复杂的模型就越有效果，而resnet、densenet这样经典模型的结构本身也并不复杂，目前的大模型也是基于Transformer模型本身具有的scalability特性加上一些实验和数据上的技巧来实现的。

个人对于一些使用特别复杂数学技巧的模型的看法是，其非常精细化地调整了特征空间，使之能与数据的分布空间更加符合，但是也正是因为如此，其泛化性能反而不够强。可能未来可以通过对特征空间的物理性可视化来进行进一步的理解和研究。

# 请专家评委批评指正

答 辩 人：邹刘文

2023/04/16