DOI: 10.3969/j.issn.2097-583X.2026.01.008

文章编号：2097-583X(2026)01-0058-08

基于特征选择算法的DBN-SVM胃癌生存期分类方法

刘道华*，余长鸣，周秋菊，王秋岱

（信阳师范大学 计算机与信息技术学院，河南 信阳 464000）

摘 要：为降低数据集的维度，筛选最优特征子集以提高胃癌预后生存期分类的准确率，提出一种融合特征选择算法的深度置信网络-支持向量机混合模型。该模型在过滤式特征选择的基础上，引入距离系数以调整整体偏移度，减少权重计算的不稳定性，从而构建新的样本权重值。在此基础上，通过Pearson相关系数分析，筛选出对胃癌生存期具有显著影响的特征子集；采用深度置信网络的受限玻尔兹曼机模块，对隐藏层中的特征子集进行特征提取；采用支持向量机，对深度置信网络的最终输出进行分类，以实现胃癌生存期的预测。通过对特征选择算法进行改进，并融合深度置信网络和支持向量机的优势，与传统单一的机器学习方法相比，该模型展现出更优的性能，其分类准确率、AUC值及$F1$值分别达到81.2%、83.4%和81.5%。

关键词：深度置信网络；支持向量机；过滤式特征选择算法；特征提取；胃癌生存期

中图分类号：TP391.4　　**文献标志码：**A　　**开放科学(资源服务)标识码(OSID)：**

A feature selection algorithm based on DBN-SVM classification for gastric cancer survival

LIU Daohua*, YU Changming, ZHOU Qiuju, WANG Qiudai

(College of Computer and Information Technology, Xinyang Normal University, Xinyang 464000, China)

Abstract: In order to reduce the dimensionality of the dataset to obtain the best feature subset as well as to improve the accuracy of the prognostic survival classification of gastric cancer, a hybrid network model of deep belief network and support vector machine combined with feature selection algorithm was proposed. Based on the filtered feature selection algorithm, a distance coefficient was introduced to adjust the overall degree of bias and reduce the instability of the calculated weight values, so as to construct new sample weight values, and then analyze the subset of features that have a greater impact on the survival period of gastric cancer through the Pearson's correlation coefficient; The constrained Boltzmann machine module was adopted in the deep belief network, and then the subset of features in the hidden layer was subjected to the feature extraction; Finally, the support vector machine was used to classify the output values of the last layer of the deep belief network to realize the classification of gastric cancer survival. By improving the feature selection algorithm and combining the advantages of deep belief network and support vector machine, the model showed better accuracy, AUC value and $F1$ value in the experiments, which are 81.2%, 83.4% and 81.5%, respectively, compared with the traditional single machine learning method.

Key words: deep belief networks; support vector machines; filtered feature selection algorithm; feature extraction; gastric cancer survival

收稿日期：2024-06-20；**修回日期：**2024-08-09；*.**通信联系人，**E-mail: ldhzzx@163.com

基金项目：国家自然科学基金项目(31872704)；河南省科技攻关项目(222102210265)；河南省本科高校研究性教学改革项目(2022SYJXLX061)；河南省高等学校重点科研项目(22A520007)；河南省研究生教育改革与质量提升工程项目(YJS2024AL104)

作者简介：刘道华(1974—)，男，河南信阳人，教授，博士，主要从事智能系统的设计及开发。

引用格式：刘道华，余长鸣，周秋菊，等.基于特征选择算法的DBN-SVM胃癌生存期分类方法[J].信阳师范大学学报(自然科学版)，2026，39(1)：58-65.

LIU Daohua, YU Changming, ZHOU Qiuju, et al. A feature selection algorithm based on DBN-SVM classification for gastric cancer survival[J].Journal of Xinyang Normal University (Natural Science Edition), 2026, 39(1):58-65.

0 引言

根据世界卫生组织2020年全球癌症发病率与死亡率数据,胃癌是死亡率和复发率较高的恶性肿瘤,分别占所有恶性肿瘤发病的5.6%和7.7%[1]。胃癌不仅缩短患者的生存时间,还加重患者家庭的经济与心理负担。目前而言,手术是胃癌患者获得根治的唯一途径。临床上通常将5年作为危险阶段的评估指标,5年内未复发的患者意味着有望临床治愈[2]。然而在当前临床诊断过程中,医生难以精准判断胃癌患者生存状况,主要依赖临床经验,这可能存在一定的风险。因此,构建胃癌患者生存期的预测模型,有助于医生制定更精准的治疗方案。

电子病历的普及与个性化医疗的兴起,为研究者们提供了大量的癌症数据,其中机器学习(Machine Learning)方法成为医学研究构建癌症预后生存期模型的重要工具[3]。常紫薇等[4]采用cox比例风险回归模型及最小化绝对收缩和选择算子(least absolute shrinkage and selection operator,LASSO)回归算法对胃癌组织与癌旁组织差异表达lncRNA进行有效分析。邓定文等[5]采用传统回归方法(单因素Cox回归)筛选出与胃癌预后相关的IncRNA,再使用机器学习迭代LASSO回归模型来预测患者的预后情况。上述方法在面对高维度数据时,仍存在噪声干扰,因此在特征选择方面存在一定的局限性。

为解决该问题,研究人员提出了一系列的方案,其中AFRESH等[6]提出使用机器学习中的ReliefF、Boruta等4种特征选择算法来筛选出最优的特征子集,然后分别传送到3个分类器:XGBoost(eXtreme Gradient Boosting)、历史梯度增强和支持向量机(Support Vector Machine,SVM)中,最终实现5年生存期的预测。孟朋辉等[7]使用改进的ReliefF与ACO特征选择算法在高维特征基因数据集上进行筛选,通过SVM分类模型对心肌病进行诊断,但此分类方法效果不佳。

近年来,深度学习和临床诊治信息深度融合。在面对大数据、高维度等复杂数据时,深度置信网络(Deep Belief Network,DBN)能够更高效地学习。BURUGADDA等[8]采用深度置信网络模型鉴别乳腺良/恶性病变,准确率达到89%。SMOLANDER等[9]结合DBN与SVM实现乳腺癌和炎症性肠病患者的诊断。以上研究表明在医学领域应用DBN的可行性和有效性。基于此,DBN和SVM相结合可作为提升胃癌生存期分类准确度的有效方法。

为剔除样本中的冗余特征,更有效地提高胃癌生存期预测模型的准确率,本文提出一种基于ED-ReliefF(Euclidean Distance-ReliefF)算法的DBN与SVM混合预测模型。一方面,采用ED-ReliefF算法进行特征选择,利用其样本间的距离系数和特征之间的差异度来更新样本权重,选择出最优特征子集;另一方面,使用DBN对筛选后的数据进行深度特征提取,利用SVM处理低维数据分类,将DBN模型与SVM分类器结合,进一步提高胃癌生存期的分类性能。

1 改进的ReliefF算法

ReliefF是一种过滤式(Filter)特征权重算法(Feature weighting algorithms)[10],它根据基因值在邻近样本间的区分程度,筛选出代表性的特征变量,并减少特征冗余,提升特征分类的精度。采用ED-ReliefF算法评估胃癌生存期特征的权重,根据权重值生成新的特征子集。

从训练集中随机选出一个样本f_i,然后从f_i同类的样本集中找出f_i的k个近邻样本f_s,从每个f_i不同类的样本集中找出k个近邻样本f_d,最后更新每个特征的权重W:

$$W(A)=W(A_0)-\mathrm{dis}(A,f_i,f_s)/(m\times k)+\sum_{c\neq \mathrm{class}(f_i)}\frac{p(c)}{1-p(\mathrm{class}(f_i))}\sum_{i=1}^{k}\frac{\mathrm{dis}(A,f_i,f_d)}{m\times k},\tag{1}$$

式中:$W(A)$表示更新后的权重值,$W(A_0)$表示初始权重值,A为特征子集,A_0表示原始数据集的基因集,m表示迭代次数,k表示近邻样本数量,f_d表示不同类别c中的第j个最近邻样本,$p(c)$为该类别的比例,$p(\mathrm{class}(f_i))$为随机选取的某样本类别的比例。

样本f_i与特征子集A中同类别的样本之间的距离为:

$$\mathrm{dis}(A,f_i,f_s)=\sum_{i=1}^{k}\frac{|f_i-\bar{f}_s|}{\max(A)-\min(A)},\tag{2}$$

式中:f_s表示与f_i属于同一类别的样本,$\bar{f}_s$表示有k个近邻样本的平均距离,$\max(A)$表示特征子集A中最大的特征值,$\min(A)$表示特征子集A中最小的特征值。样本f_i与特征子集A中不同类别样

本间的距离：

$$\mathrm{dis}(A,f_i,M_j(c))=\sum_{c\neq \mathrm{class}(f_i)}\frac{p(c)}{1-p(\mathrm{class}(f_i))}\times\sum_{i=1}^{k}\frac{|f_i-\overline{M_j(c)}|}{\max(A)-\min(A)},\tag{3}$$

式中：$p(c)$表示目标样本c与总样本的概率比，$p(\mathrm{class}(f_i))$表示f_i的样本概率比，$\overline{M_j(c)}$表示k个近邻样本间的平均距离。

通过该样本中f_i计算出个体样本与群体样本平均基因表达之间的总偏移量R：

$$R=\sqrt{\sum_{i=1}^{k}(f_i-\bar{f})}/k,\tag{4}$$

式中：$\bar{f}$为所选样本特征的平均值。

以上参数构建的距离系数，如式(6)所示。

$$\partial=1\times10^{-10},\tag{5}$$

$$\mathrm{CD}=R/\sum_{i=1}^{k}f_i+\partial,\tag{6}$$

式(5)中：∂代表避免除零错误。

两个样本特征差异程度越大，则特征差距越明显，距离系数也就越大，式(6)与式(1)相结合可以降低计算的不稳定性。最终样本f_i在进行特征权重计算时，算法中特征权重系数的更新式(7)为：

$$W(A)=W(A_0)-\mathrm{CD}\times\mathrm{dis}(A,f_i,f_s)/(m\times k)+\sum_{c\neq \mathrm{class}(f_i)}\frac{p(c)}{1-p(\mathrm{class}(f_i))}\times\sum_{i=1}^{k}\frac{\mathrm{dis}(A,f_i,M_j(c))}{m\times k}\times\mathrm{CD}。\tag{7}$$

2　基于ED-ReliefF算法的DBN-SVM网络模型

2.1　深度置信网络原理

在DBN中采用了多层受限玻尔兹曼机(Restricted Boltzmann Machine，RBM)方法，其中RBM可作为特征提取器，包含可见层和隐藏层[11]，如图1所示。

在RBM模型中，定义了能量函数$E(v,h|\theta)$，如式(8)所示。

$$E(v,h|\theta)=-\sum_{i=1}^{m}\sum_{j=1}^{n}w_{ij}v_ih_j-\sum_{i=1}^{m}a_iv_i-\sum_{j=1}^{n}b_jh_j,\tag{8}$$

式中：w_{ij}表示隐藏层中的第i个神经元与可见层中的第j个神经元的权重值，a_i表示隐藏层第i个神经元的偏置，b_j表示可见层第j个神经元的偏置，m表示隐藏层的数量，n表示可见层的数量。

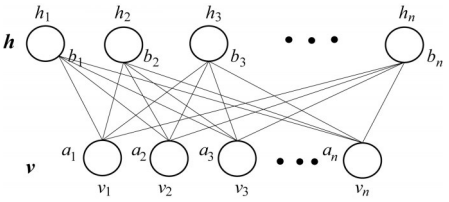

图1　RBM模型

Fig. 1　RBM model

根据式(8)得出状态(v,h)的联合概率密度，其定义为式(9)，归一化分子Z如式(10)所示。

$$p(v,h|\theta)=\frac{1}{Z}\mathrm{e}^{-E(v,h|\theta)},\tag{9}$$

$$Z=\sum_{i=1}^{m}\sum_{j=1}^{n}\mathrm{e}^{-E(v,h|\theta)}。\tag{10}$$

隐藏层中第j个单元的被激活概率可表示为：

$$p(n_j=1|v,\theta)=f(\sum_{j=1}^{m}w_{ij}v_i+a_j)。\tag{11}$$

同理，可见层中的第i个单元的被激活概率可表示为：

$$p(v_i=1|v,\theta)=f(\sum_{j=1}^{m}w_{ij}h_j+b_i)。\tag{12}$$

在RBM训练过程中，采用对比散度算法(Contrastive Divergence，CD)[12]得到参数集(w_{ij},a_i,b_j)，各参数的更新准则如式(13)、(14)和(15)所示。

$$\Delta w_{ij}=\varepsilon(\langle v_ih_j\rangle_{\mathrm{data}}-\langle v_ih_j\rangle_{\mathrm{recon}}),\tag{13}$$

$$\Delta a_i=\varepsilon(\langle v_i\rangle_{\mathrm{data}}-\langle v_i\rangle_{\mathrm{recon}}),\tag{14}$$

$$\Delta b_j=\varepsilon(\langle h_j\rangle_{\mathrm{data}}-\langle v_j\rangle_{\mathrm{recon}})。\tag{15}$$

式中：ε是学习率，$\langle\cdot\rangle_{\mathrm{data}}$表示训练数据的期望，$\langle\cdot\rangle_{\mathrm{recon}}$表示重构后模型定义的分布。

2.2　支持向量机分类器

DBN完成预训练和微调后，需要对胃癌生存期进行预测，采用SVM对胃癌生存期进行诊断。SVM在高维空间构建最佳分离超平面，以分离不同的决策类别。模型训练过程中，SVM可以成功分离训练样本，具体条件见式(16)。

$$y_i(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i+\boldsymbol{b})-1\geqslant0,\tag{16}$$

式中：$\boldsymbol{b}$代表偏置，$\boldsymbol{\omega}$代表权重向量，$\boldsymbol{x}_i$代表输入的训练集数据。

为了找到最优的分离超平面，并且使间隔距离最大化，可使用凸二次规划来获得最佳分离超平面，如式(17)所示。

$$\min \frac{1}{2}\| \boldsymbol{\omega} \|^2。 \tag{17}$$

同时，对于凸二次规划问题需要解出α_i^*来确定最优超平面的$\boldsymbol{\omega}^*$参数和$\boldsymbol{b}^*$，所以最优超平面决策函数可表示为：

$$f(\boldsymbol{x})=\mathrm{sign}(\boldsymbol{\omega}^{*\mathrm{T}}\boldsymbol{x}_i+\boldsymbol{b}^*)。 \tag{18}$$

对于非线性可分样本，通常使用核函数来构建分类器。通过引入非松弛变量ξ和惩罚函数C来解决低维空间中的线性不可分问题，如式(19)所示。

$$\min \frac{1}{2}\| \boldsymbol{\omega} \|^2+C\sum_{i=1}^{N}\xi_i。 \tag{19}$$

所以新的最优超平面决策函数为：

$$f(\boldsymbol{x})=\mathrm{sign}(\sum_{i=1}^{N}a_i^* y_i k(\boldsymbol{x}\cdot\boldsymbol{x}_i)+b^*), \tag{20}$$

式中：$k(\boldsymbol{x}\cdot\boldsymbol{x}_i)$表示核函数。

采用径向基函数(RBF)，如式(21)所示。

$$k(\boldsymbol{x},\boldsymbol{x}_i)=\exp(-\gamma\| \boldsymbol{x}-\boldsymbol{x}_i\|), \tag{21}$$

式中：$\boldsymbol{x}$是输入向量，$\boldsymbol{x}_i$是中心向量，$\| \boldsymbol{x}-\boldsymbol{x}_i\|$是它们之间的欧氏距离，$\gamma$是单调函数。

为了优化SVM分类器的性能，使用网格搜索法，在超参数的候选空间中进行搜索，从而找到最佳的C和γ组合。在搜索的同时，使用十倍交叉验证法[13]对每个超参数组合进行评估，以确保所选择的参数组合在不同的数据子集上都具有良好的泛化性能。

2.3 基于ED-ReliefF算法的DBN-SVM分类方法

胃癌分期预测由三部分构成：特征选择、特征提取和分类。为减少数据集中的冗余特征、噪声和不平衡性，引入ED-ReliefF算法对原始数据进行特征选择，通过式(7)得到全新的权重值，采用皮尔逊相关系数热力图分析对胃癌分期预测影响显著的特征子集。图2展示了胃癌数据中各特征的相关关系热力图。

由图2可知，T、M、N、肿瘤直径、肿瘤分化程度和原发灶手术信息关联性强，进而形成新的数据集，随后对数据进行特征提取，具体分类过程如图3所示。

DBN能够提取特征子集中更深层的信息，为了提高DBN在分类方面的能力，需要与SVM分类器结合。在DBN中逐层训练RBMs，为学习更

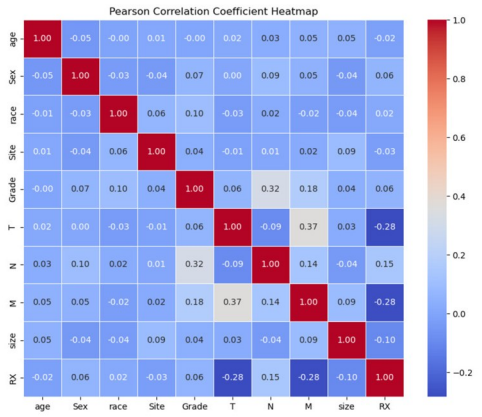

图2 特征相关分析热力图

Fig. 2 Heat map of feature correlation analysis

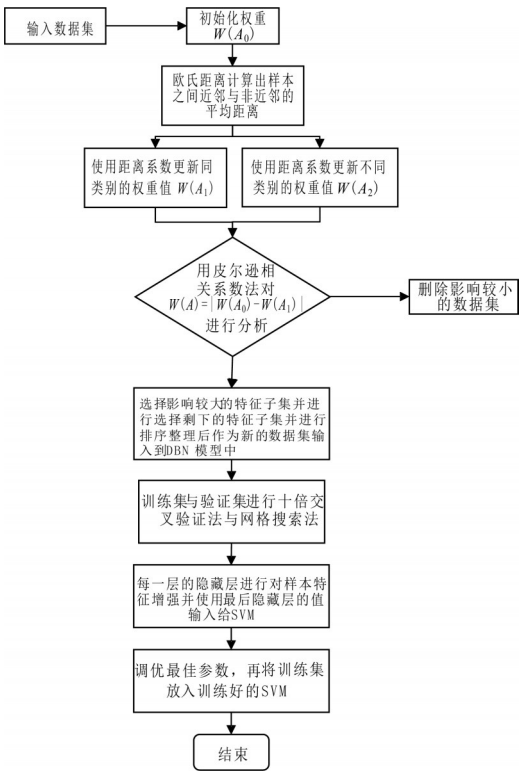

图3 分类流程图

Fig. 3 Classification flowchart

高层次的抽象特征，除最后一层添加激活函数tanh外，其他每层添加ReLU激活函数，以避免模型过拟合。采用反向传播算法迭代调整DBN中每个连接的权重，此过程优化了DBN特征的提取能力。再利用SVM分类器对提取的深层特征进行分类。由此构建的DBN-SVM模型如图4所示。

基于ED-ReliefF算法的DBN-SVM模型计算具体步骤如下：

Input：输入训练集train dataset，测试集test dataset，验证集valid dataset，训练集对应的标签

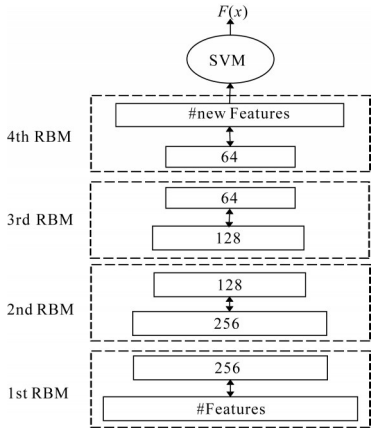

图4　DBN-SVM模型结构图

Fig. 4　Structure diagram of DBN-SVM model

train labels,迭代次数N, k个最近邻样本。

Output:生存期的标签为$F(x)$。

1. 初始化式(1)中的权重$W(A)$;

2. 输入处理后的train dataset,使用式(2)获得k个最近邻样本,若样本f_i与特征子集A中的样本为不同类别,则使用式(3);

3. 迭代循环N次,使用式(6)计算所有特征子集的距离系数;

4. 使用式(7)来更新权值;

5. 使用皮尔逊相关系数分析出对胃癌分期影响显著的特征子集。选取最优的特征子集进行排序整理并作为新的train dataset;

6. 将train dataset、valid dataset分别放入DBN模型;

7. 其中每个样本映射到模型中,各隐藏层对样本特征进行增强并以第i个隐藏层的输出作为x_i_train、x_i_valid;

8. x_i_train和train labels 训练SVM;

9. 使用网格搜索法和十倍交叉验证法进行调参,选择最优SVM模型;

10. 将valid dataset放入训练好的SVM模型,从而得到预测测试标签$F(x)$。

3　实验仿真与分析

3.1　数据集及预处理

胃癌疾病数据来源于seer数据库17 Registries,Nov 2022 sub(2000—2020)里面的数据(https://seer.cancer.gov/data/),包括11项临床特征:诊断年龄、性别、种族、肿瘤直径、肿瘤原发灶、肿瘤分级、T、N、M、原发部位手术信息和生存时间等11个变量(10个特征属性、1个标签属性)。数据集将诊断年份设置在2010—2015年;根据美国癌症联合委员会肿瘤分期系统第六版(American Joint Committee on Cancer, AJCC 6th Edition)将肿瘤部位选择胃(stomach);病理学类型(Histologic Type)选择腺癌(8140—8389);删除了一些噪声,例如:Tx、T1NO、T4NOS、N3NOS、NX、NA和blanks等,确保数据的准确性,排除手术、放疗、化疗等信息未知的病例;排除仅有尸检或死亡证明的病例。

采用多重插补法来处理缺失值,以确保对缺失数据的充分补充。利用Z-score技术进行数据缩放,使用SMOTE[14]算法,以提高模型的准确性和鲁棒性。最终收集到17 584个样本,并随机分为3部分:训练集9403例、测试集4150例和验证集4031例。

3.2　实验参数设置

为了评估胃癌预后5年生存率预测模型的有效性,在配置为IntelCore i5-7200、8 GB内存、Windows操作系统的设备上进行实验,采用改进的ReliefF特征选择算法筛选出最优的特征子集,利用DBN进行有效特征的提取,再用SVM测试模型的有效性。为验证所提模型的性能,与文献[15-17]中的模型进行了对比。采用Python语言的TensorFlow库构建SVM、XGBoost分类器[18]、随机森林(Random Forest,RF)、K近最邻近分类算法(K-Nearest Neighbor, KNN)、多层感知机(Multilayer perceptron,MLP)等5种方法与本文模型进行比较,观察不同模型的分类效果。各个模型的参数设置如表1所示。

3.3　评价指标

在癌症生存期研究分析中,越来越多的研究人员采用不同的分类器对癌症患者生存期进行预测。为了验证SVM、RF、XGBoost、KNN、MLP和本文模型DBN-SVM等6种模型在预测胃癌生存期中的性能,采用准确率(Accuracy)、$F1$值($F1$-Score)、召回率(Recall)、AUC等作为评估模型的评价指标。

3.4　实验结果及分析

为了确保数据验证的一致性和可靠性,采用预处理后的数据样本,并对每种分类算法进行网格搜索和十倍交叉验证。如图5所示,在验证集结果中,SVM的AUC值为0.733,RF的AUC值为0.746,而XGBoost的AUC值为0.755。可见,

表1 各个模型参数

Tab. 1 Various model parameters

序号	模型	参数
1	SVM	C='20',gamma:'0.001',kernel='rbf'
2	XGBoost	Learning_rate:'0.1',max_depth='3',n_estimators:'100',colsample_bytree:'0.8'
3	RF	Verbose:'2',Max_depth:'10',n_estimator:'100',random_state'='888'
4	KNN	K=5
5	MLP	Learning_rate:'0.001',hidden_layer_size='(128,64,32)',max_iter:'50'
6	DBN-SVM	Activation1='relu',Activation2='relu',Activation3='relu',Activation4='tanh',optimizer='adam', C='20',gamma:'0.001',kernel='rbf'

SVM在整个假阳率(False Positive Rate, FPR)范围内的性能较弱,特别是在FPR中间区域,其真阳性率(True Positive Rate, TPR)提升较慢。相比之下,XGBoost的表现略优于RF和SVM。由图5可知,DBN-SVM模型的AUC值低于MLP模型的AUC值。

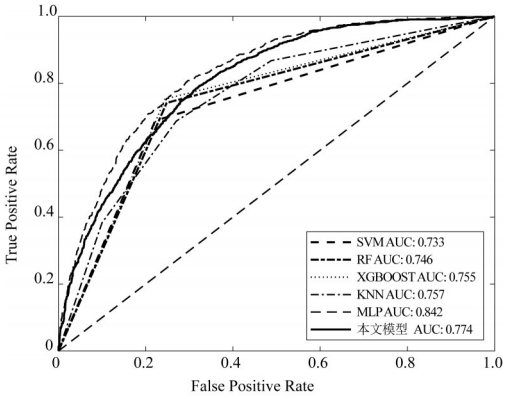

图5 6种模型在验证集AUC值对比图

Fig. 5 Comparison of AUC values of the six models in the validation set

结合图6可见,MLP的AUC值与DBN-SVM的AUC值虽较为接近,但DBN-SVM模型在整个FPR范围内都高于其他模型,DBN-SVM的AUC值为0.834。同时,在测试集的结果中SVM的AUC值为0.727,RF的AUC值为0.764,XGBoost的AUC值为0.777,SVM曲线整体较低,尤其是在FPR较低的区域,其TPR增长较慢,RF的性能优于SVM。从图6可知,RF相对于XGBoost略差,KNN曲线在FPR较低的区域上升较快,但在FPR高中高段稍显平缓,性能不及MLP模型好。

下面比较6种不同预测模型的性能。这些模型均使用筛选出的特征子集进行训练,目的是评估它们预测胃癌患者生存期的效果,如表2所示。

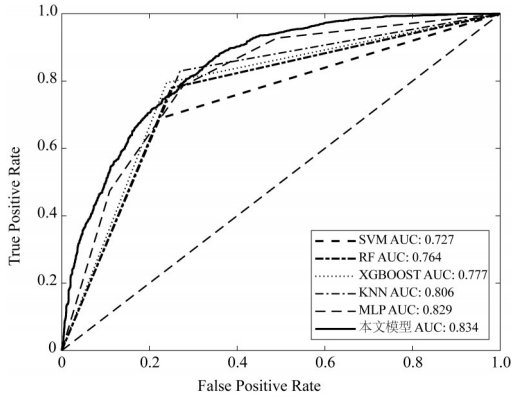

图6 6种模型在测试集AUC值对比图

Fig. 6 Comparison of AUC values of the six models in the test set

表2 6种模型预测效果对比表 %

Tab. 2 Comparison of the predictive effects of the six models

模型	Accuracy	Recall	Specificity	*F*1-Score
SVM	72.6	67.1	78.1	70.8
XGBoost	77.7	79.5	76.0	77.8
RF	76.5	77.9	75.0	76.5
KNN	75.7	80.1	71.4	76.4
MLP	75.1	75.8	75.2	75.2
DBN-SVM	81.2	80.5	79.1	81.5

从综合评测指标Accuracy、Recall、Specifity、*F*1-Score值来看,DBN-SVM的Accuracy值达到81.2%,Recall值达到80.5%,Specificity值达到79.1%,*F*1值达到81.5%,均高于其余5个模型。在Recall方面,与KNN模型相比,DBN-SVM模型的表现一般,但在筛选最优特征子集后,通过DBN增强关键特征权重,并与SVM结合,显著提升了分类准确率。为综合评估6种分类器在胃癌患者生存期预测中的表现,绘制了各模型评价指标的柱状图,6种模型的性能对比结果如图7所示。

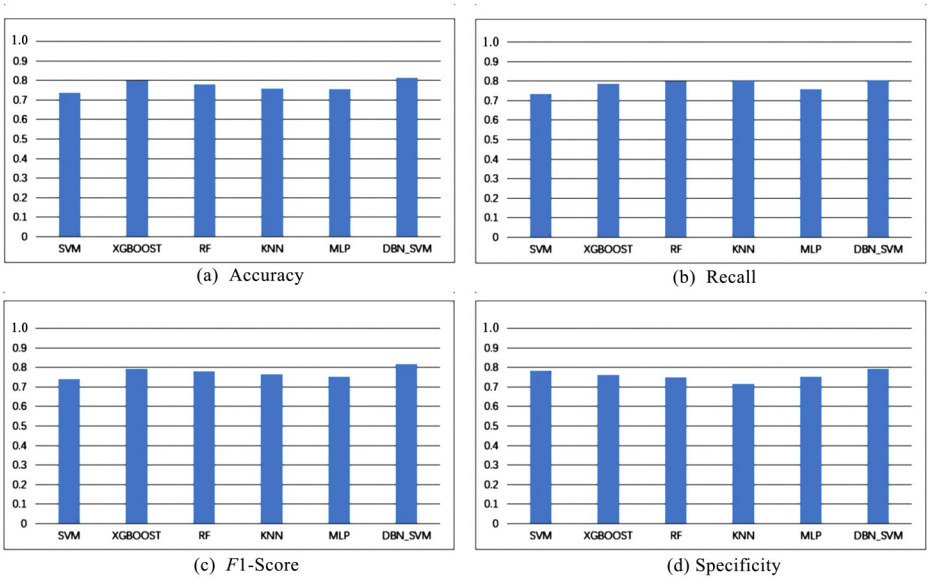

图7　6种不同模型的分类器的评价指标对比图

Fig. 7　Comparison of evaluation metrics for classifiers using six different models

本文提出的DBN-SVM模型在AUC指标和整体的评价指标方面较其他5种模型具有优势,在胃癌患者生存期预测方面也表现出良好的性能。

4　结束语

提出一种基于改变权重值分析的最优子集ED-ReliefF算法与DBN-SVM混合分类模型,对比传统的机器学习分类器和深度学习的神经网络模型,可以较好地解决准确率低等问题,更好地捕捉特征之间的关系。通过DBN模型可以提取深层数据中的隐藏特征,再与SVM分类器结合输出最终结果。

本文提出的方法较单一机器学习分类器效果更优,提高了胃癌生存期预测的准确率,可为医务人员对患者制定合适的治疗方案提供依据和帮助。其他深度学习模型的特征提取与分类能力与DBN-SVM模型的对比,将是下一步的研究工作。

参考文献:

[1] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: A Cancer Journal for Clinicians, 2021, 71(3): 209-249.

[2] 高美虹, 尚学群. 利用人工智能预测癌症的易感性、复发性和生存期[J]. 生物化学与生物物理进展, 2022, 49(9): 1687-1702.
GAO Meihong, SHANG Xuequn. Artificial intelligence based prediction for cancer susceptibility, recurrence and survival[J]. Progress in Biochemistry and Biophysics, 2022, 49(9): 1687-1702.

[3] 陈雯, 王旭, 段辉宏, 等. 深度学习在癌症预后预测模型中的应用研究[J]. 生物医学工程学杂志, 2020, 37(5): 918-929.
CHEN Wen, WANG Xu, DUAN Huihong, et al. Application of deep learning in cancer prognosis prediction model [J]. Journal of Biomedical Engineering, 2020, 37(5): 918-929.

[4] 常紫薇, 刘辉, 张秋萌, 等. 基于TCGA和LASSO回归的胃癌预后lncRNA预测模型构建[J]. 临床肿瘤学杂志, 2020, 25(9): 823-829.
CHANG Ziwei, LIU Hui, ZHANG Qiumeng, et al. Establishment of lncRNA predictive model based on TCGA and LASSO for prognosis of gastric cancer[J]. Chinese Clinical Oncology, 2020, 25(9): 823-829.

[5] 邓定文, 殷中强. 构建胃癌lncRNA预测模型以促进胃癌的治疗[J]. 临床医学进展, 2023, 13(12): 19461-19470.
DENG Dingwen, YIN Zhongqiang. Constructing lncRNA prediction model of gastric cancer to promote the treatment of gastric cancer[J]. Advances in Clinical Medicine, 2023, 13(12): 19461-19470.

[6] AFRASH M R, MIRBAGHERI E, MASHOUFI M, et al. Optimizing prognostic factors of five year survival in gastric cancer patients using feature selection techniques with machine learning algorithms: A comparative study[J].

BMC Medical Informatics and Decision Making, 2023, 23(1): 54.

[7] 孟朋辉, 黄凯雯, 徐磊. 基于改进ReliefF与ACO特征选择算法的心肌病分类模型[J]. 软件工程与应用, 2022, 11(2): 267-281.
MENG Penghui, HUANG Kaiwen, XU Lei. Optimized feature selection algorithm based on ReliefF and ant colony for cardiomyopathy classification[J]. Software Engineering and Applications, 2022, 11(2): 267-281.

[8] BURUGADDA V R, PATIL V C, VYAS N, et al. Enhancing breast cancer diagnosis: A deep belief network approach for mammography image analysis [C]//2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT), Faridabad, 2023: 46-51.

[9] SMOLANDER J, DEHMER M, EMMERT-STREIB F. Comparing deep belief networks with support vector machines for classifying gene expression data from complex disorders[J]. FEBS Open Bio, 2019, 9(7): 1232-1248.

[10] 吴辰文, 李晨阳, 郭叔瑾, 等. 基于ReliefF和蚁群算法的特征基因选择方法[J]. 计算机应用研究, 2018, 35(9): 2610-2613.
WU Chenwen, LI Chenyang, GUO Shujin, et al. Feature gene selection method based on ReliefF and ant colony optimization[J]. Application Research of Computers, 2018, 35(9): 2610-2613.

[11] CHAO Hao, SONG Cheng, LU Baoyun, et al. Feature extraction based on DBN-SVM for tone recognition[J]. Journal of Information Processing Systems, 2019, 15(1): 91-99.

[12] SU Xiyuan, CAO Changqing, ZENG Xiaodong, et al. Application of DBN and GWO-SVM in analog circuit fault diagnosis[J]. Scientific Reports, 2021, 11(1): 7969.

[13] 高鹏. 基于改进深度置信网络的胃癌诊断预测模型研究与应用[D]. 合肥: 安徽大学, 2021.
GAO Peng. Research and application of gastric cancer diagnosis prediction model based on improved deep belief network[D]. Hefei: Anhui University, 2021.

[14] MOHAMMED A J, HASSSAN M M, KADIR D H. Improving classification performance for a novel imbalanced medical dataset using SMOTE method [J]. International Journal of Advanced Trends in Computer Science and Engineering, 2020, 9(3): 3161-3172.

[15] AFRASH M R, SHAFIEE M, KAZEMI-ARPANAHI H. Establishing machine learning models to predict the early risk of gastric cancer based on lifestyle factors[J]. BMC Gastroenterology, 2023, 23(1): 6.

[16] TIAN Huakai, LIU Zitao, LIU Jiang, et al. Application of machine learning algorithm in predicting distant metastasis of T1 gastric cancer[J]. Scientific Reports, 2023, 13(1): 5741.

[17] DURKAYA KURTCAN B, OZCAN T. Predicting customer churn using grey wolf optimization-based support vector machine with principal component analysis[J]. Journal of Forecasting, 2023, 42(6): 1329-1340.

[18] MA Baoshan, YAN Ge, CHAI Bingjie, et al. XGBLC: An improved survival prediction model based on XGBoost [J]. Bioinformatics, 2022, 38(2): 410-418.

责任编辑:郭红建　陈松楠