Journal of the Korea Industrial Information Systems Research Vol. 21 No.6, Dec. 2016 :13-21
http://dx.doi.org./10.9723/jksiis.2016.21.6.013 ISSN:1229-3741

Neighborhood 러프집합 모델을 활용한 유방 종양의 진단적 특징 선택

(A Diagnostic Feature Subset Selection of Breast Tumor Based on Neighborhood Rough Set Model)

손 창 식[1)*], 최 락 현[2)], 강 원 석[3)*], 이 종 하[4)]
(Chang-Sik Son, Rock-Hyun Choi, Won-Seok Kang, and Jong-Ha Lee)

요 약 특징선택은 데이터 마이닝, 기계학습 분야에서 가장 중요한 이슈 중 하나로, 원본 데이터에서 가장 좋은 분류 성능을 보여줄 수 있는 특징들을 찾아내는 방법이다. 본 논문에서는 정보 입자성을 기반으로 한 neighborhood 러프집합 모델을 이용한 특징선택 방법을 제안한다. 제안된 방법의 효과성은 5,252명의 유방 초음파 영상으로부터 추출된 298가지의 특징들 중에서 유방 종양의 진단과 관련된 유용한 특징들을 선택하는 문제에 적용되었다. 실험결과 19가지의 진단적 특징을 찾을 수 있었고, 이때에 평균 분류 정확성은 97.6%를 보였다.

핵심주제어 : Neighborhood 러프 집합, Neighborhood 근사화, 특징 선택, 유방 종양 진단

Abstract Feature selection is the one of important issue in the field of data mining and machine learning. It is the technique to find a subset of features which provides the best classification performance, from the source data. We propose a feature subset selection method using the neighborhood rough set model based on information granularity. To demonstrate the effectiveness of proposed method, it was applied to select the useful features associated with breast tumor diagnosis of 298 shape features extracted from 5,252 breast ultrasound images, which include 2,745 benign and 2,507 malignant cases. Experimental results showed that 19 diagnostic features were strong predictors of breast cancer diagnosis and then average classification accuracy was 97.6%.

Key Words : Neighborhood Rough Set, Neighborhood Approximations, Feature Selection, Breast Tumor Diagnosis

* Corresponding Author : changsikson@dgist.ac.kr, wskang@dgist.ac.kr
† 본 연구는 산업통상자원부에서 지원하는 산업핵심기술개발사업(10063553)에 의해 수행되었습니다.
Manuscript received Nov, 24, 2016 / revised Dec, 08, 2016 / accepted Dec, 12, 2016
1) DGIST 웰니스융합연구센터, 제1저자, 교신저자
2) DGIST 웰니스융합연구센터, 제2저자
3) DGIST 웰니스융합연구센터, 제3저자, 교신저자
4) 계명대학교 의과대학 의용공학과, 제4저자

1. 서 론

유방암은 여성에게 가장 흔한 암으로 알려져 있고, 유방암의 사망률은 점차적으로 감소하고 있으나 발생률은 급격히 증가하고 있다[1]. 2011년에 보고된 중앙암등록본부 자료에 의하면, 유방암은 2009년 국내 여성의 암 발생률 중에서 갑상선암에 이어 2위를 차지하고 있으며, 지난 10년간 유방암 환자 수는 약 2.5배로 증가한 것으로 조사되었다. 발생연령은 40대 환자에서 약 37.1%로 가장 많은 분포를 나타내고, 한 해에 약 1만 명씩 발생하는 것으로 조사되었다[2]. 또한 유방암 수술환자의 5년 생존율은 1기는 99%, 2기는 89%에 이르나 3기와 4기의 경우 59%에서 28%로 급격히 떨어진다[3].

일반적으로 임상에서는 유방암의 조기 발견과 진단을 위해서 주로 3가지 방법, 유방 자가검진 방법, 유방의 정기적 진찰 방법(clinical breast examination), 그리고 영상 의학적 검진법을 병행할 것을 권장하고 있다. 특히 영상의학적 검진법에서 활용되는 대표적인 유방암 진단방법으로는 유방 촬영술(mammography), 유방 초음파술(breast ultrasonography), 유방 MRI을 활용하고 있다[3]. 이 중 선별적 진단 도구로서 가장 많이 활용되는 방법으로는 유방 촬영술이며, 76-94%의 민감도(sensitivity)와 90% 이상의 특이도(specificity)를 보인다[4]. 하지만 선별적 유방 촬영술은 방사선 피폭에 따른 암 발생 위험이 증가할 뿐만 아니라, 위음성(false negative)과 위양성(false positive) 진단에 따른 불필요한 추가 검진, 조직검사 등의 단점으로 인하여 득보다는 실을 고려하여 결정할 것을 권장하고 있다[5].

최근에는 이러한 진단적 오류 등을 최소화하기 위해서 유방 촬영술과 유방 초음파술을 병행한 다양한 연구결과들이 보고되고 있으며, 미국방사선의학회(the american college of radiology imaging network)에서 수행된 임상시험연구 ACRIN 6666가 대표적인 예이다[6-8]. ACRIN 6666은 유방암 고위험군 여성에서 유방 초음파검진의 효과성을 살펴본 연구로서, 유방 촬영술만을 시행한 경우보다 유방 초음파술을 병행할 경우 1,000명 당 4.2명의 유방암 환자를 더 발견할 수 있으며, 조직검사판정에 대한 양성예측도(positive predictive value, PPV)는 유방 촬영술만 사용된 경우 22.6%, 유방 초음파술만 사용된 경우 8.9%, 이들 검사를 병행할 경우 11.2% 증가된 것으로 보고하고 있다. 그리고 유방 초음파술의 경우 방사선 피폭이 없고, 유방 초음파에서 발견된 병변에 대한 조직검사가 간단하여 추가적인 진단적 도구로 많이 활용되고 있다[9-11]. 그러나 유방 초음파술은 사용된 측정장비에서 제공된 영상의 품질이 시행자의 진단결과에 직접적인 영향을 줄 수 있으므로, 이를 개선하기 위한 다양한 컴퓨터 보조진단(computer-aided diagnosis, 이하 CADx) 지원기술들이 필요하다.

대부분의 유방암 진단을 위한 CADx 기술들은 크게 특징추출(feature extraction), 특징선택(feature selection), 및 분류(classification)로 구성되어 있으며, 이들 단계를 상호 결합하여 유방 종양의 양성(benignancy)/악성(malignancy)여부를 구별하는데 응용되고 있다[12]. 특히 특징추출 과정은 CADx 기술에서 가장 중요한 전처리 과정으로, 주로 영상의 품질 개선기법[13], 형태학적 특성분석을 위한 영상분할기법[14-17] 등의 영상신호처리 기술들이 활용된다. 그리고 특징선택 과정은 수집된 전체 특징들 가운데에 유방 종양을 구별하는데 효과적인 후보 특징들을 선별하는데 주로 사용되며, 퍼지 클러스터링(fuzzy clustering), 벡터 양자화기(learning vector quantization, LVQ), 인공 면역 시스템(artificial immune system, AIS), 인공 신경망(artificial neural network, ANN), 서포트 벡터 머신(support vector machine, SVM) 등의 방법들이 주로 활용되고 있다[18-21]. 이들 방법들 중에서 F-score 특징분석기법을 결합한 SVM[21] 방법이 가장 좋은 분류 정확성(99.51%)을 보여준다. 하지만 SVM의 경우 많은 수의 데이터를 학습해야 하는 경우 오랜 학습시간이 소요되고, 기대하는 수준의 분류성능을 얻지 못할 가능성이 있으며, 결측치(missing value), 이상치(outlier)와 같은 잡음(noise)을 포함한 데이터의 경우 결측치 보상(imputation) 혹은 제거와 같은 추가적인 전처리 과정이 필요하다.

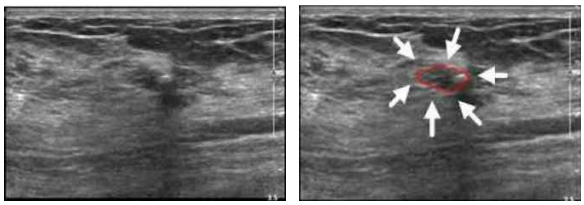
본 연구에서는 정보의 입자성(granularity) 개

념을 활용한 neighborhood 러프집합 모델을 소개하고, 이를 활용한 특징선택 방법을 제안한다. 제안된 방법은 5,252명의 유방 초음파 영상으로부터 추출된 298가지의 진단적 특징 데이터를 이용하여 악성종양을 가진 환자를 식별하는 문제에 적용함으로써 그 효과성을 보이고자 한다.

2. 연구방법

2.1 유방 초음파 영상획득 및 데이터 수집

2006년에서 2010년 사이, 삼성의료원 의학연구윤리심의위원회의(institutional review board of Samsung Medical Center)의 연구 승인 후 5,252명의 유방암 의심환자의 유방 초음파 영상(Philips ATL iU222 초음파 기계, 스캐너 5-12MHz, 6cm linear probe 활용)을 획득하였다. 유방 초음파 영상에서 의학적 소견 상 양성종양(benign tumor)을 가진 환자는 2,745명으로 조사되었으며, 연령은 11세에서 81세의 분포를 보였으며, 악성종양(malignant tumor)을 가진 환자는 2,507명으로 연령은 24세에서 86세의 분포를 보였다.



(a) (b)

Fig. 1 The Example of Detection of a Breast Tumor[17]. (a) The Original Image with a Benign Tumor, (b) Manual Segmentation of the Tumor

유방 종양의 관심영역(region of interest, ROI)은 방사선 전문의에 의해서 수작업으로 표시되었고 (Fig. 1(b)), 종양의 37가지의 형태학적 특성들은 Lee[17]의 방법을 이용하여 290가지의 진단적 특징들을 추출하였고[22], 이외에 종양의 히스토그램 특성 등과 관련된 8가지 진단적 특징들을 추가하였다(Table 1).

Table 1 Diagnostic Features Extracted from Breast Ultrasound Images

Extractor	Feature No.	Feature Name
1	F1-F140	Spatial gray-level dependence matrix(SGLD)
2	F141-F203	Fourier with shape context
3	F204-F234	Fourier with centroid distance(Magnitude)
4	F235-F265	Fourier with centroid distance(Phase)
5	F266	Intensity in the mass area
6	F267	Gradient magnitude in the mass area
7	F268	Orientation
8	F269	Depth-width ratio
9-11	F270-F272	Distance between mass shape and best fit ellipse
12-17	F273-F278	The average gray changes between tissue area and mass area
18	F279	The average gray changes between posterior and mass area
19-22	F280-F283	The histogram changes between tissue and mass(bin 0-3)
23	F284	Compare the gray value of left, post and right under lesion
24	F285	The number of lobulate areas
25	F286	The number of protuberances
26	F287	The number of depressions
27	F288	Lobulation index
28-29	F289-F290	Elliptic-normalized circumference
30-35	F291-F296	Histogram (mean, variance, skewness, kurtosis, energy, entropy)
36-37	F297-F298	Fourier power spectrum (annual-ring and wedge sampling geometries)

2.2 Neighborhood 러프집합 모델을 활용한 특징선택

일반적인 정보시스템(information system)은 $IS=<U,A>$로 정의된다. U는 샘플들의 집합(universe) $U=\{x_1,x_2,...,x_n\}$, A는 속성 (혹은 특징)들의 집합 $A=\{a_1,a_2,...,a_m\}$을 의미한다. 속성들의 집합 A가 $A=C\cup D$로 구분될 때, 정보시스템 IS는 $DS=<U,C\cup D>$로 정의가 가능하며, 이를 결정시스템(decision system)이라 부른다. C는 m개의 독립변수로 구성된 조건속성(condition attributes)의 집합, D는 하나 이상의 종속변수로 구성된 결정속성(decision attribute)이고, 각 샘플 데이터 x_i는 S개의 클래스 중 하나 $x_i\in c_t$, $t=1,2,...,S$로 분류된다.

[정의 1] 특징 공간(feature space) B 상에서 i-번째 샘플 x_i의 이웃(neighborhood) $\delta(x_i)$, $x_i\in U$, $B\subseteq C$는 다음과 같이 정의된다[23-24].

$$\delta_B(x_i)=\{x_j|x_j\in U,\triangle^B(x_i,x_j)\le\delta\} \tag{1}$$

x_i와 x_j는 i-와 j-번째 샘플 데이터, $\triangle^B(x_i,x_j)$는 특징 공간 B 상에서 두 샘플 데이터의 유사도(similarity degree)를 나타낸 거리척도이고, δ, $\delta\in[0,1]$은 이웃 반경(neighborhood radius)을 결정하는 기준을 나타낸다. 또한 $\triangle$는 다음의 3가지 함수적 성질을 만족한다.

i) Reflexive, $\triangle(x_i,x_j)\ge 0$, $\triangle(x_i,x_j)=0$ if $x_i=x_j$
ii) Symmetric, $\triangle(x_i,x_j)=\triangle(x_j,x_i)$
iii) Transitive, $\triangle(x_i,x_k)\le\triangle(x_i,x_j)+\triangle(x_j,x_k)$.

이들 샘플 데이터 간에 유사도 평가는 식 (2)의 Minkowsky 척도를 활용하며, P의 값에 따라 3가지 평가함수, Manhattan (P=1), Euclidean (P=2), 및 Chebychev($P=\infty$)로 정의가 가능하다.

$$\triangle_P(x_i,x_j)=\left(\sum_{k=1}^{m}|f(x_i,a_k)-f(x_j,a_k)|^P\right)^{1/P} \tag{2}$$

식 (2)에서 $f(x_i,a_k)$는 k-번째 조건속성에서 i-번째 샘플 데이터의 값을 나타낸다.

[정의 2] 모든 샘플들의 이웃 입자성(neighborhood granularity)은 식 (3)을 이용하여 관계 행렬, $N=(r_{ij})_{n\times n}$로 표현이 가능하다[23-24].

$$r_{ij}=\begin{cases}1, \text{if}\ \triangle(x_i,x_j)\le\delta,\\ 0, otherwise.\end{cases} \tag{3}$$

N에서 r_{ij}는 i-와 j-번째 샘플 데이터 간에 이진관계(binary relation)이고, i-번째 샘플 데이터에 대한 이웃 입자성은 $\delta(x_i)=\bigcup_{j=1}^{n} r_{ij}$로 표현이 가능하다. 또한 이는 이웃 반경 δ의 범위 내에 존재하는 샘플들의 부분집합(subset)을 의미한다. 하지만 식 (3)은 특정 샘플 데이터의 속성에서 결측치를 포함할 경우, 해당 샘플의 이웃 입자성을 구성할 수 없으므로, 결측치 보상/제거를 위한 추가적인 전처리 과정이 필요하다. 따라서 본 연구에서는 식 (4)와 같이 관계 행렬을 수정하였고, '*'는 결측치, '∧'는 논리 AND 연산자를 나타낸다.

$$r_{ij}^{*}=\begin{cases}1, \text{if}\ f(x_i,a_k)\ne *\wedge f(x_j,a_k)\ne *\wedge\triangle(x_i,x_j)\le\delta,\\ k=1,2,...,m,\\ 0, otherwise.\end{cases} \tag{4}$$

Fig. 2는 위에서 정의된 이웃 입자성에 관한 하나의 예로서, 2차원 Euclidean 공간에서 이웃 반경 $\delta_i(i=1,...,4)$가 0.2에서 1.0으로 변화는 동안에 샘플 x_1에 대한 입자성의 변화를 보여준다.

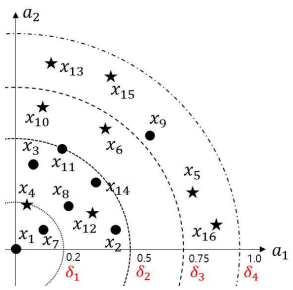


Fig. 2 The Variation of Granules According to Neighborhood Radius

그림에서 각 샘플은 2개의 클래스(●와 ★)로 구분된다고 가정한다. 만약 이웃 반경 δ값이 δ_1 (δ_1=0.2)로 정의된다면, 샘플 x_1의 이웃 입자성으로 3개의 샘플 $\{x_1,x_4,x_7\}$을 고려하게 된다. 이때 해당 이웃 입자성 내 샘플 x_4의 클래스는 나머지 두 개의 샘플이 포함된 클래스와 비교해 볼 때 다르다. 또한 이웃 반경 δ값이 δ_3(δ_3=0.75)로 변경된다면, x_1의 이웃 입자성은 11개의 샘플을 포함하게 되고, 이들 중 4개의 샘플 $\{x_4,x_6,x_{10},x_{12}\}$은 해당 샘플과 서로 다른 클래스 특성을 보여준다.

일반적으로 이웃 반경 δ값이 증가할수록 이웃 입자성에 포함된 샘플 수는 증가하는 반면에, 샘플들 간 클래스 특성을 고려한 유사성은 낮아지는 경향을 보인다. 그러므로 이웃 반경 δ값 혹은 범위를 찾는 것은 정보 입자성을 이용한 특징 선택방법에서 중요한 문제 중 하나이다. Hu[24]는 이웃 반경 δ값에 대한 최적의 범위를 찾기 위해서, 다양한 특징들로 구성된 혼합속성 데이터에서 이웃 반경 δ의 영향에 대한 실험을 수행하였고, [0.1,0.3] 범위에서 결정될 때 보다 좋은 분류 성능을 제공할 수 있다는 결과를 보고하였다.

[정의 3] 이웃 근사 공간(neighborhood approximation space)은 전체 샘플 데이터의 집합 U와 이를 통해 유도된 관계 행렬 N으로부터 $<U,N>$으로 표현이 가능하다[23-24]. 이때 결정속성 D 상에 정의된 개념 (즉 동일한 클래스에 포함된 샘플들의 부분집합)을 활용하여 하한(lower)과 상한(upper) 근사는 다음의 식을 이용하여 구성할 수 있다.

$$\underline{N}X=\{x_i|\delta(x_i)\subseteq X,x_i\in U\} \quad (5)$$

$$\overline{N}X=\{x_i|\delta(x_i)\cap X\neq\varnothing,x_i\subseteq U\} \quad (6)$$

하한 근사 $\underline{N}X$는 i-번째 이웃 입자성 $\delta(x_i)$가 결정속성 D상에 정의된 임의의 개념 X에 완전히 포함되는 경우를, 반면 상한 근사 $\overline{N}X$는 $\delta(x_i)$가 부분적으로 포함되는 경우를 나타내며, $\underline{N}X\subseteq X\subseteq\overline{N}X$의 특성을 만족한다.

[정의 4] 이웃 근사 공간에서 임의의 조건속성 혹은 속성의 부분집합에 대한 의존도(dependency degree)는 식 (7)로부터 계산될 수 있다[23-24].

$$\gamma_B(D)=\frac{|\cup\underline{N}X_k|}{|U|} \quad (7)$$

Table 2 Feature Subset Selection Algorithm

```
Input: Decision System or Decision Table <U,C∪D>.
Output: Feature Subset R.
R←∅
do
    T←R
    ∀ a_k∈(C− T)
    for k = 1 to m
        Calculate the dependency degree γ_{a_k}(D)
    end
    Find a feature a_k^* with the highest dependency degree arg_k max{γ_{a_k}(D)}
    if γ_{R∪{a_k^*}}(D) > γ_T(D)
        R ← {a_k^*}
    end
until γ_R(D) = γ_C(D)
return R
```

|.|은 해당 조건을 만족하는 샘플들의 수, $\gamma_B(D)$는 특징 공간 B, $B \in C$에서 정의된 이웃 입자성으로부터 결정속성 D에서 정의된 개념들을 근사적으로 추정할 수 있는 정도를 나타낸다. 만약 $\gamma_B(D)=1$이라면, 결정속성 D는 B로서 완전히 해석 가능함을 의미하고, $0<\gamma_B(D)<1$이면 부분적으로 해석이 가능하다.

하지만 전체 조건속성의 집합 C에서 $\gamma_B(D)=1$인 조건을 만족하는 속성의 부분집합을 찾는 것은 비결정 난해(non-deterministic polynomial time-hard, HP-hard) 문제로 잘 알려져 있다. 따라서 본 연구에서는 전방향 선택 전략(forward selection strategy)을 기반으로 한 Hill Climbing 탐색법을 이용하여 부분 최적해를 찾고자 하였고, Table 2는 이를 위한 의사코드(pseudo code)를 나타낸다.

특징 선택 알고리즘은 초기에 $R=\varnothing$으로 초기화하고, 식 (7)의 평가함수를 이용하여 각 조건속성에 대한 의존도를 순차적으로 평가하고, 이들 중 가장 큰 의존도를 가진 특징을 선택 한다. 만약 선택된 조건속성의 의존도가 $\gamma_T(D)$의 의존도보다 큰 경우 후보 특징으로 선택하고, 이 과정은 의존도의 차이가 없을 때까지 반복하게 된다.

3. 실험결과

실험에서는 Euclidean 거리척도($P=2$)를 이용하여 이웃 반경 δ값을 0.05에서 0.20 사이에서 0.05 간격으로 변화시켰을 때, 유방 종양의 선택된 진단적 특징들의 변화와 이때에 분류 정확성을 평가하였다. 좀 더 객관적인 성능평가 실험을 위해서, 통계적 10-겹 교차검정(10-fold cross validation; train, 90%; test, 10%) 실험을 수행하였고, 로지스틱 회귀분석 모델을 이용하여 실험 데이터의 출력을 예측하였다. 또한 교차검정 실험동안에 분류 성능은 Confusion Matrix를 기반으로 한 수신자 조작 특성(receiver operating characteristic, ROC) 곡선의 AUC (area under the ROC) 평가척도를 이용하였다[25-26].

실험에서 사용된 모든 알고리즘들은 Windows 10 Professional X64, Intel(R) Core(TM) i7-6700 @3.40GHz, RAM 32GB, JavaSE-1.7 (JDK1.7.0_45), Eclipse IDE for Java Developers <version: Luna Service Release 1(4.4.1)> 환경에서 개발되었다.

Table 3은 이웃 반경 δ값을 조절하였을 때, 유방 종양의 선택된 진단적 특징들과 평균 분류 정확성, 그리고 소요된 시간을 나타낸다. 'Before Feature Subset Selection'은 전체 샘플 데이터를 이용하여 10-겹 교차검정 실험을 수행한 경우를, 'After Feature Subset Selection'은 제안된 특징 선택방법을 적용한 후 10-겹 교차검정 실험을 수행한 경우를 의미한다. 'CPU Time'은 특징 선택과정에서 소요된 시간과 로지스틱 회귀분석 모

Table 3 Average Classification Accuracies According to Neighborhood Radius

Performance Evaluation	Neighborhood Radius(δ)	CPU Time(sec.)		Avg. AUC
		Feature Selection	Classifier	
Before Feature Subset Selection	-	-	1981.024	0.934
After Feature Subset Selection	0.05[1)]	93.829	1.124	0.961
	0.10[2)]	119.431	2.316	0.970
	0.15[3)]	136.881	2.845	0.974
	0.20[4)]	197.702	2.832	0.976

1), {F226, F297, F3, F204, F97, F224, F26}

2), {F226, F202, F232, F108, F3, F204, F69, F279, F228, F214, F4}

3), {F225, F219, F112, F286, F204, F111, F280, F203, F69, F295, F134, F14, F205, F132}

4), {F128, F135, F226, F232, F231, F286, F14, F204, F138, F275, F279, F295, F83, F3, F205, F212, F283, F111, F13}

델(즉 분류기)에서 소요된 시간을 구분하였다.

실험결과, Table 1에서 제시된 전체 298가지의 유방종양의 진단적 특징을 사용한 경우에 비해, 제안된 특징 선택방법을 적용한 경우의 실험에서 보다 좋은 분류 정확성을 보여주었고, 분류기에서 소요된 시간 또한 개선되었다. 실험에서 가변적으로 조절된 이웃 반경 δ값의 경우, 0.20일 때 가장 좋은 평균 AUC 97.6%를 보여 주었고, 이 때에 선택된 유방 종양의 진단적 특징들로는 'Spatial Gray-level Dependency Matrix(SGLD)'와 관련된 특징들 {F3, F13, F14, F83, F111, F128, F135, F138}, 'Centroid Distance for Magnitude'와 관련된 특징들 {F204, F205, F212, F226, F231, F232}, 'Average Gray Changes Between Tissue Area and Mass Area'와 관련된 특징들 {F275, F279}, 'Histogram Changes Between Tissue and Mass'와 관련된 특징 {F283}, 'Number of Lobulate Areas'와 관련된 특징 {F286}, 'Histogram Entropy'와 관련된 특징 {F295}인 것으로 조사되었다.

또한 위의 결과는 선행연구[17,22]의 결과 (즉 분류기 SVM with radial basis kernel function 활용)에서 제시된 각각의 평균 분류 정확성 92.0%와 95.0%와 비교해 볼 때, 평균 5.6%와 2.6% 개선된 결과를 보여준다.

4. 결 론

본 논문에서는 neighborhood 러프집합 모델을 기반으로 한 특징선택 알고리즘을 제안하였다. 제안된 알고리즘의 효과성을 평가하기 위해서 유방 초음파 영상으로부터 추출된 유방 종양의 진단적 특징들 가운데에 유방암을 진단하는 데에 가장 효과적인 후보 특징들을 분석하였다. 298가지의 진단적 특징들 중에서 'Spatial Gray-level Dependency Matrix(SGLD)', 'Centroid Distance for Magnitude', 'Average Gray Changes Between Tissue Area and Mass Area', 'Histogram Changes Between Tissue and Mass', 'Number of Lobulate Areas', 'Histogram Entropy'와 관련된 19가지 진단적 특징들이 유방 종양을 구별하는데 중요한 특징들로 선택되었으며, 평균 97.6%의 분류 정확성을 보여주었다.

향후 연구에서는 수집된 데이터로부터 이웃 반경 δ값을 자동 결정할 수 있는 방법에 대한 연구와 다양한 특징 선택방법들과의 비교 연구를 수행할 계획이다.

References

[1] F. Bessaoud, J. P. Daures, "Dietary Factors and Breast Cancer Risk: A Case Control Study Among a Population in Southern France," Nutrition and Cancer, Vol. 60, No. 2, pp. 177-187, 2008.

[2] Yoo, Y. G., Choi, S. K., Hwang, S. J., and Kim, H. S., "Risk Factors of Breast Cancer According to Life Style," Journal of The Korea Contents Association, Vol. 13, No. 4, pp. 262-272, 2013.

[3] Nam, S. J., "Screening and Diagnosis for Breast Cancers," Journal of The Korean Medical Association, Vol. 52, No. 10, pp. 946-951, 2013.

[4] Chung, S. Y. and Han. B. K., "Breast Diagnostic Imaging", Seoul, Ilchokak, 2006.

[5] Bae, J. M., "On the Benefits and Harms of Mammography for Breast Cancer Screening in Korean Woman," Korean Journal of Family Practice, Vol. 4, No. 1, pp. 1-6, 2014.

[6] Berg, W. A., "Rationale for a Trial of Screening Breast Ultrasound: American College of Radiology Imaging Network (ACRIN) 6666," American Journal of Roentgenology, Vol. 180, No. 5, pp. 1225-1228, 2003.

[7] Berg, W. A., "Supplemental Screening Sonography in Dense Breasts," Radiologic Clinics of North America, Vol. 42, No. 5, pp. 845-851, 2004.

[8] Berg, W. A., "Beyond Standard

Mammographic Screening: Mammography at Age Extremes, Ultrasound, and MR Imaging," Radiologic Clinics of North America, Vol. 45, No. 5, pp. 895-906, 2007.

[9] Lim, S. Y., Lee, S. J., Shin, Y. K., Lee, S. N., Choi, J. Y., and Kang, D. R., "Comparison of the Diagnostic Value Between Mammography and Mammography with Breast Ultrasonography in Diagnosing Breast Cancer," Korean Journal of Family Medicine, Vol. 24, No. 10, pp. 925-933, 2003.

[10] Berg, W. A., Blume, J. D., and Cormack, J. B., "Combined Screening with Ultrasound and Mammography vs. Mammography Alone in Women at Elevated Risk of Breast Cancer," JAMA, Vol. 299, No. 18, pp. 2151-2163, 2008.

[11] Jang, J. M., Yi, A., and Koo, H. R., "Differentiation of Breast Mass Using Automated Breast US: Application of US BI-RADS lexicon," Journal of The Korean Society for Breast Screening, Vol. 8, No. 2, pp. 115-120, 2011.

[12] Giger, M. L., Chan, H. P., and Boone, J. "Anniversary Paper: History and Status of CAD and Quantitative Image Analysis: The Role of Medical Physics and AAPM," Medical Physics, Vol. 35, No. 12, pp. 5799-5820, 2008.

[13] Guo, Y. H., Cheng, H. D., Huang, J. H., Tai, J. W., Zhao, W., and Sun, L., "Breast Ultrasound Image Enhancement Using Fuzzy Logic," Ultrasound in Medicine and Biology, Vol. 32, No. 2, pp. 237-247, 2006.

[14] Cheng, J. Z., Chou, Y. H., C. S., Huang, Chang, Y. C., Tiu, C. M., and Chen, K. W., "Computer-Aided US Diagnosis of Breast Lesions by Using Cell-Based Contour Grouping," Radiology, Vol. 255, No. 3 pp. 746-754, 2010.

[15] Sahiner, B., "Computer-Aided Characterization of Mammographic Masses: Accuracy of Mass Segmentation and its Effects on Characterization," IEEE Transactions on Medical Imaging, Vol. 20, No. 12, pp. 1275-1284, 2001.

[16] Joo, S., Yang, Y. S., Moon, W. K., and Kim, H. C., "Computer-Aided Diagnosis of Solid Breast Nodules: Use of an Artificial Neural Network Based on Multiple Sonographic Feature," IEEE Transactions on Medical Imaging, Vol. 23, No. 10, pp. 1292-1300, 2004.

[17] Lee, J. H., Seong, Y. K., Chang, C. H., Park, J. M., Park, M. H., and Woo, K. G., "Fourier-Based Shape Feature Extraction Technique for Computer-Aided B-Mode Ultrasound Diagnosis of Breast Tumor," 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), August 28-September 1, San Diego, California, USA, pp. 6551-6554, 2012.

[18] Abonyi, J. and Szeifert, F., "Supervised Fuzzy Clustering for the Identification of Fuzzy Classifiers," Pattern Recognition Letters, Vol. 14, No. 24, pp. 2195-2207, 2003.

[19] Goodman, D. E., Boggess, L., and Watkins, A. "Artificial Immune System Classification of Multiple-Class Problems," In Proceedings of Intelligent Engineering Systems, pp. 179-184, 2002.

[20] Setiono, R., "Generating Concise and Accurate Classification Rules for Breast Cancer Diagnosis," Artificial Intelligence in Medicine, Vol 18, No.3, pp. 205-217, 2000.

[21] Akay, M. F., "Support Vector Machines Combined with Feature Selection for Breast Cancer Diagnosis," Expert Systems with Applications, Vol. 36, No. 2, pp. 3240-3247, 2009.

[22] Lee, J. H., Seong, Y. K., Chang, C. H., Ko, E. Y., Cho, B. H., and Ku, J. H.,

"Computer-Aided Lesion Diagnosis in B-Mode Ultrasound by Border Irregularity and Multiple Sonographic Features," Proceedings of SPIE, Medical Imaging 2013, February 2013.

[23] Hu, Q. H., Yu, D. R., and Xie, Z. X., "Neighborhood Classifiers," Expert Systems with Applications, Vol. 34, No. 2, pp. 866-876, 2008.

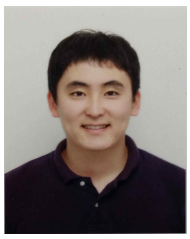
[24] Hu, Q. H., Yu, D. R., Liu, J. F., and Wu, C. X., "Neighborhood Rough Set Based Heterogeneous Feature Subset Selection," Information Sciences, Vol. 178, No. 18, pp. 3577-3594, 2008.

[25] Son, C. S., Kang, W. S., Choi, R. H., Park, H. S., Han, S. W., and Kim, Y. N., "A Probabilistic Knowledge Model for Analyzing Heart Rate Variability," Journal of the Korea Industrial Systems Research, Vol. 20, No. 3, pp. 61-69, 2015.

[26] Ha, S. H. and Zhang, Z. Y., "Empirical Evaluation of Ensemble Approach for Diagnostic Knowledge Management," The Journal of Information Systems, Vol. 20, No. 3, pp. 237-255, 2011.

손 창 식 (Chang-Sik Son)

- 정회원
- 대구가톨릭대학교 전자정보공학부 공학사
- 대구가톨릭대학교 전산통계학과 이학석사
- 대구가톨릭대학교 전산통계학과 공학박사
- DGIST 웰니스융합연구센터 선임연구원
- 관심분야 : 인공지능, 기계학습, 빅데이터 마이닝, 유헬스 및 웰니스 플랫폼



최 락 현 (Rock-Hyun Choi)

- 정회원
- 대구대학교 통신공학과 공학사
- 대구대학교 정보통신학과 공학석사
- DGIST 웰니스융합연구센터 연구원
- 관심분야 : 데이터마이닝, WNCS, 임베디드 시스템, CPS

강 원 석 (Won-Seok Kang)

- 정회원
- 영남대학교 컴퓨터공학과 공학사
- 영남대학교 컴퓨터공학과 공학석사
- DGIST 웰니스융합연구센터 선임연구원
- 관심분야 : 의용생체데이터 처리, 기계학습, 데이터마이닝, BCI/BMI 분산병렬 시뮬레이션 및 모델링

이 종 하 (Jong-Ha Lee)

- 정회원
- 인하대학교 전자공학과 공학사
- New York University Polytechnic Institute 전기컴퓨터공학과 공학석사
- Temple University 전기컴퓨터공학과 공학박사
- 계명대하교 의과대학 의용공학과 교수
- 관심분야 : 인공지능, 컴퓨터진단, 초음파영상분석, 생체신호처리