

컴퓨터 단층 촬영 영상에서의 전이성 척추 종양의 정량적 분류를 위한 라디오믹스 기반의 머신러닝 기법

이은우^{1,2} · 임상현^{2,3} · 전지수² · 강혜원^{1,2} · 김영재^{1,2} · 전지영^{4*} · 김광기^{1,2,3*}

¹가천대학교 보건과학대학 의용생체공학과, ²가천대학교 의과대학 의공학과교실

³가천융합의과학원 융합의과학과, ⁴가천대학교 길병원 영상의학과

Radiomics-based Machine Learning Approach for Quantitative Classification of Spinal Metastases in Computed Tomography

Eun Woo Lee^{1,2}, Sang Heon Lim^{2,3}, Ji Soo Jeon², Hye Won Kang^{1,2}, Young Jae Kim^{1,2},
Ji Young Jeon^{4*} and Kwang Gi Kim^{1,2,3*}

¹Department of Biomedical Engineering, College of Health Science, Gachon University

²Department of Biomedical Engineering, Gil Medical Center, College of Medicine, Gachon University

³Department of Health Sciences & Technology, Gachon Advanced Institute for Health Sciences & Technology (GAIHST)

⁴Department of Radiology, Gil Medical Center, Gachon University College of Medicine

(Manuscript received 11 March 2021 ; revised 4 June 2021 ; accepted 18 June 2021)

71

Abstract: Currently, the naked eyes-based diagnosis of bone metastases on CT images relies on qualitative assessment. For this reason, there is a great need for a state-of-the-art approach that can assess and follow-up the bone metastases with quantitative biomarker. Radiomics can be used as a biomarker for objective lesion assessment by extracting quantitative numerical values from digital medical images. In this study, therefore, we evaluated the clinical applicability of non-invasive and objective bone metastases computer-aided diagnosis using radiomics-based biomarkers in CT. We employed a total of 21 approaches consist of three-classifiers and seven-feature selection methods to predict bone metastases and select biomarkers. We extracted three-dimensional features from the CT that three groups consisted of osteoblastic, osteolytic, and normal-healthy vertebral bodies. For evaluation, we compared the prediction results of the classifiers with the medical staff's diagnosis results. As a result of the three-class-classification performance evaluation, we demonstrated that the combination of the random forest classifier and the sequential backward selection feature selection approach reached AUC of 0.74 on average. Moreover, we confirmed that 90-percentile, kurtosis, and energy were the features that contributed high in the classification of bone metastases in this approach. We expect that selected quantitative features will be helpful as biomarkers in improving the patient's survival and quality of life.

Key words: Radiomics, Machine learning, Spinal metastases, Quantitative biomarker, Computer-aided diagnosis

*Corresponding Author : Kwang Gi Kim

Department of Biomedical Engineering, Gil Medical Center, Gachon University College of Medicine, 21, Namdong-daero 774 beon-gil, Namdong-gu, Incheon, 21565, Republic of Korea.

Tel: +82-32-458-2770

E-mail: kimkg@gachon.ac.kr

*Corresponding Author : Ji Young Jeon

Department of Radiology, Gachon University College of Medicine, Gil Medical Center, 38-13 Docjeom-ro 3-bungil, Namdong-gu, Incheon, 21565, Korea

Tel: +82-10-8435-8035

E-mail: mdjeonjy@gilhospital.com

This research was supported by the Gil Medical Center (FRD2019-11-02), and by the GRRC program of Gyeonggi Province (No. GRRC Gachon 2020-B01).

I. 서 론

골 전이암(bone metastasis)은 유방, 전립선, 갑상선, 폐 및 신장에서 발생하는 악성 종양이 뼈로 전이되는 병변이다 [1,2]. 임상적으로 골 전이암은 척추와 같은 중심 골격부위에 호 발되는 것으로 알려져 있다 [3,4]. 원발병소로부터 척추에 암 이 전이된 환자들은 척추 압박과 병리학적 골절의 통증을 느끼며, 추적 관찰 결과 골 전이 이후의 1년 생존율은 약 20%, 중앙생존기간은 5-6개월로 보고된 바 있다 [5,6]. 또한, 골

전이암으로 인해 환자가 받는 고통은 환자의 삶의 질을 저하시키며, 부작용의 발생을 일으킬 수 있다. 따라서, 골 전이암의 조기진단은 골 전이암이 환자들에게 미치는 영향을 조기에 최소화하여 골 전이 부작용들의 발생을 예방한다는 점과 환자의 삶의 질 향상과 다소 간의 생명연장이 가능하다는 면에서 필요하다[7]. 골 전이암에 의한 증상을 추적하는 과정에서 암의 원발병소를 발견한 기존 보고에 의하면, 골 전이암의 조기진단은 발견하지 못한 원발병소를 추적할 수 있다는 점에서도 중요하다[8].

골 전이암은 골 주사(bone scintigraphy)와 같이 방사성 동위원소를 이용한 핵의학적 검사를 시행할 경우 정밀한 진단이 가능하나, computed tomography(CT)영상 분석을 통해 연조직의 병변뿐만 아니라 골 전이암에 대한 평가가 가능하다[9]. 따라서 대부분의 정밀진단은 이전의 CT검사결과에 따라 추가 검사 여부가 결정되며, 현재 골 전이암의 조기진단은 CT영상을 사용하여 수행되고 있다. CT영상을 이용한 골 전이암의 진단은 병소로 판단되는 영역의 단면영상을 비침습적으로 검사하는 것이 가능하다는 장점이 있으나, 진단의 정확성이 의료진의 임상적 경험에 크게 의존한다는 한계가 있다[10,11]. 이러한 한계점을 극복하기 위해 CT영상에서 골 전이 여부를 정량적으로 측정 및 평가할 수 있는 생체지표(biomarker)를 선정하기 위한 연구는 그 필요성이 높다. 이와 같은 생체지표는 골 전이 진단에서 핵의학적 검사와 의료진의 경험에 대한 의존도를 감소시킬 수 있으며, 골 전이의 객관적인 평가가 가능하다. 그 뿐만 아니라, 컴퓨터 보조 진단(Computer-Aided Diagnosis, CAD) 시스템 연구 및 개발에서 유용한 도구로 사용될 수 있다[12].

의학 분야에서 라디오믹스(radiomics)는 영상 데이터를 특성화하는 기법을 이용하여 디지털 의료 영상에서 다양한 특징을 추출하는 방법이다. 상기한 방법들을 이용하여 추출한 특징(feature)을 통해 육안으로 인식하기 어려운 영상학적 특성을 정량적인 값으로 표현할 수 있다. 라디오믹스 기반의 특징 추출 방법은 크게 세 가지로 분류할 수 있으며, 그 종류는 다음과 같다: (1) 영상의 히스토그램(histogram) 기반의 분석으로 얻어진 first-order 특징 값, (2) 영상의 표면 분석을 기반으로 복셀(voxel) 간의 통계적 상관관계(correlation)를 분석하는 Gray Level Co-occurrence Matrix(GLCM), Gray Level Run Length Matrix(GLRLM) 등의 행렬로부터 얻어진 second-order 특징 값, (3) wavelet, Laplacian of Gaussian(LoG)와 같이 영상에 필터링 처리를 하여 반복적이거나 비-반복적인 패턴을 추출하는 higher-order 특징 값[13]. 상기한 방법들을 이용한 라디오믹스는 CT, Magnetic Resonance Imaging(MRI), Positron Emission Tomography(PET)과 같은 디지털 의료영상에서 관찰되는 병변의 형태(shape) 또는 표면(texture)의 특징을 정량적인 수

치로 표현하기에 적합한 기법이다. 또한, 라디오믹스는 병소로 추정되는 관심영역(Region of Interest, ROI) 내부에서 다양한 종류의 정량적인 특징 값을 추출(extraction)하여 생체지표로써 사용이 가능하다는 장점이 있다.

이러한 장점에 따라, 최근 몇 년 동안 라디오믹스를 기반으로 의료영상에서 추출한 특징 값 중에서 병소의 평가에 사용될 수 있는 유의미한 특징 값을 선택한 뒤, 이를 생체지표로 이용하여 환자의 지속적인 추적 및 관찰에서 정량적인 병소평가의 가능성을 제시한 몇 가지 연구사례가 보고되었다. 기존 연구에서는 라디오믹스를 기반으로 설계한 딥러닝(deep learning) 모델이 비소세포폐암의 재발을 예측할 수 있음을 제시하였다[14]. 소개한 기존연구에서 비소세포폐암의 재발 여부를 예측한 딥러닝 모델의 경우, 암 재발 예측성능이 71.5%의 정확도까지 도달함을 증명하였다. 또한, CT영상에서 라디오믹스를 이용한 폐암 정밀 진단 및 임상 적용 가능성에 대해 조사한 연구에 따르면, 라디오믹스 기반의 CT영상 분석은 폐암의 조기예측, 수술 후 폐 기능 평가 및 이환율 예측 가능성을 제시하였다[15]. 그러나, 라디오믹스 기반으로 의료영상에서 추출한 모든 값은 수 백가지의 다양한 특징들로 이루어져 있으며, 이들 중 병소 평가에 유의한 영향을 끼치는 특정한 특징을 생체지표로 선택하는 것은 여전히 어려운 작업이다. 이러한 한계점에 따라, 데이터 마이닝(data mining)을 통한 병변평가에 유의미한 특징의 선택(feature selection) 작업이 필수적임에도 불구하고, 특정 몇 가지의 특징 값들이 병변 평가에 유의한 기여(contribution)를 하는지에 대한 조사는 부족하다.

따라서, 본 연구에서는 골 모세포성 골전이(blastic), 골 용해성 골전이(lytic), 그리고 정상적으로 건강한(normal-healthy) 추체를 정량적으로 분류할 수 있는 생체지표를 찾고자 한다. 실험에서는 특징 선택 기법을 이용하여 척추 전이 평가에 유의미한 특징을 선택한 후 분류기의 조합으로 얻어진 라디오믹스 기반 척추 전이 분류 기법의 머신러닝 골 전이암 예측 결과와 전문의의 진단 결과를 비교 분석하여 제안한 기법의 골 전이 보조 진단 기법의 성능을 평가한다. 또한, 분류기의 분류 성능을 평가함과 동시에 가장 높은 분류 성능까지 도달한 분류기에서 기여도가 높은 유의한 특징과 해당 특징의 중요도(importance)를 조사한다.

II. 연구 방법

1. CT영상 및 관심영역 데이터 수집

골 전이암은 뼈 재형성 조절 과정에서 과잉 골 형성을 유도하는 골 모세포성 골전이와 골 형성이 억제되어 뼈를 용해시키는 골 용해성 골전이로 나뉜다(그림 1abc). 실험에 사용된 데이터는 가천대학교 길병원에서 두 가지의 골 전이암과

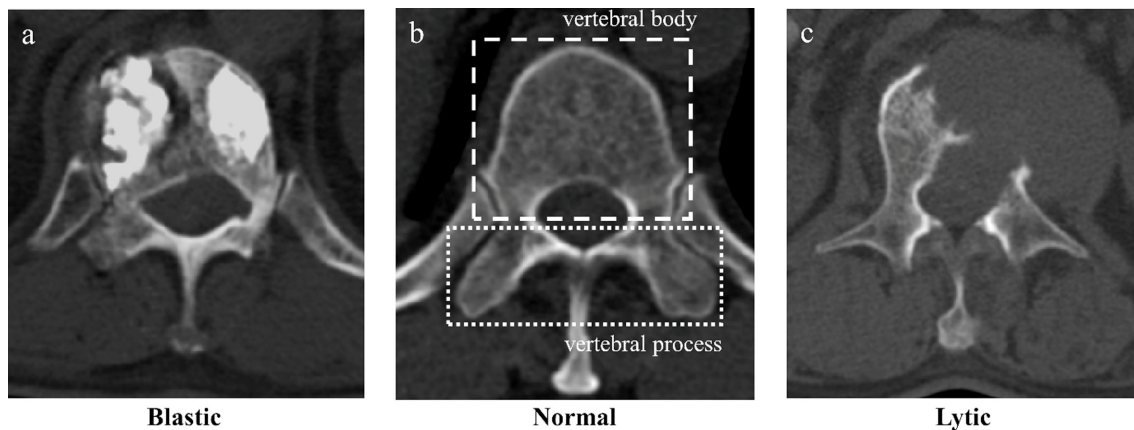


그림 1. 골 모세포성 골 전이암을 포함하는 척추(a), 정상 척추(b), 골 용해성 골전암을 포함하는 척추(c)의 그림. 척추의 추체(b, 위), 돌기 영역(b, 아래)을 구분하여 제시하였다.

Fig. 1. From the left, the blastic(a), normal(b), and lytic(c) spinal metastases were presented. In (b), we presented the vertebral body(top) and vertebral process(bottom) separately.

건강한 정상의 추체를 지닌 환자들을 대상으로 개체의 특성에 따라 30명씩 총 90명 환자의 CT 데이터를 수집하였다 (IRB No. GDIRB2019-137). 90명의 데이터는 평균 60.6 ± 15.0 세의 48명의 남성과 42명의 여성 환자로 구성되어 있다.

본 논문에서는 척추뼈를 이루고 있는 추체(vertebral body)와 돌기(vertebral process) 중 골 전이암 발병빈도가 높은 추체에서의 골 전이암 병변 검출을 목표로 한다(그림 1b). 관심영역의 수동-분할하기 위해 3D Slicer(ver.4.11, <http://download.slicer.org/>)[16]에서 3차원의 값이 잘 영상 분할 되었는지 확인하기 위해서 MPR(Multi planner reconstruction) 방법을 사용하여, 골 용해성 골전이, 골 모세포성 골전이, 정상 환자들의 추체의 영역을 각 30개씩 총 90개의 추체를 수동-분할하였다. 수동 분할된 추체의 정확성은 두 명의 영상 의학과 전문의의 확인 결과를 통해 평가받았다.

2. 라디오믹스 기반 특징 추출

육안을 이용한 시각적 영상 분석은 의료진의 숙련도와 주관적인 경험에 영향을 받으며, 정량적인 분석이 어려워 병변 진단을 객관화하는데 한계가 있다. 라디오 믹스는 영상 데이터에서 수학적, 통계적 기법을 이용하여 병변의 특성을 추출함으로써 기존의 시각적 영상분석에서 얻을 수 없는 객관적 정보를 추출해 낼 수 있다. 따라서 본 연구에서는 골 전이암의 특징을 정량적인 수치로 표현하기 위해 CT영상에서 라디오믹스 기반의 특징 추출을 수행하였다. 라디오믹스의 특징추출 방법은 총 3가지가 있으며, 본 연구에서는 영상의 히스토그램의 기반 분석으로 얻은 17개의 first order 특징과 second-order 특징에서 픽셀(pixel)간의 상관관계 정보를 추출하는 24개의 GLCM, 동일한 그레이 레벨 값을 가진 픽셀을 탐색하는 16개의 GLZLM, 픽셀 군집(cluster)

의 크기와 강도를 탐색하는 16개의 GLSZM으로 상기한 2가지의 특징 추출 기법을 이용하여 총 73개의 특징을 추출하였다[17-19].

일반적으로, 방사선 영상의 분석을 위해서는 윈도우 중앙 및 범위(window level/width) 값을 설정하여 골을 관찰하기 적합한 hounsfield unit(HU)를 가지도록 정규화(normalization) 처리를 수행한다. 그러나, 본 연구에서 사용된 추체 관심영역에서는 병변의 특성을 포함하고 있기 때문에 일반적인 골이 가지는 HU의 중앙 및 범위에서 예외가 존재한다. 따라서, 본 연구에서는 전처리 과정에서 원본 CT에서 관찰할 수 있는 병변 정보의 손실을 줄이기 위해 CT영상의 HU를 보존한 상태에서 Window width의 값과 Window level의 값은 각각 평균 453(450-500), 1446(1200-1500)을 기준으로 라디오믹스 접근법을 이용하여 척추체의 3차원 특징을 추출하였다.

3. 특징 선택 및 분류기(feature selection & classifier)

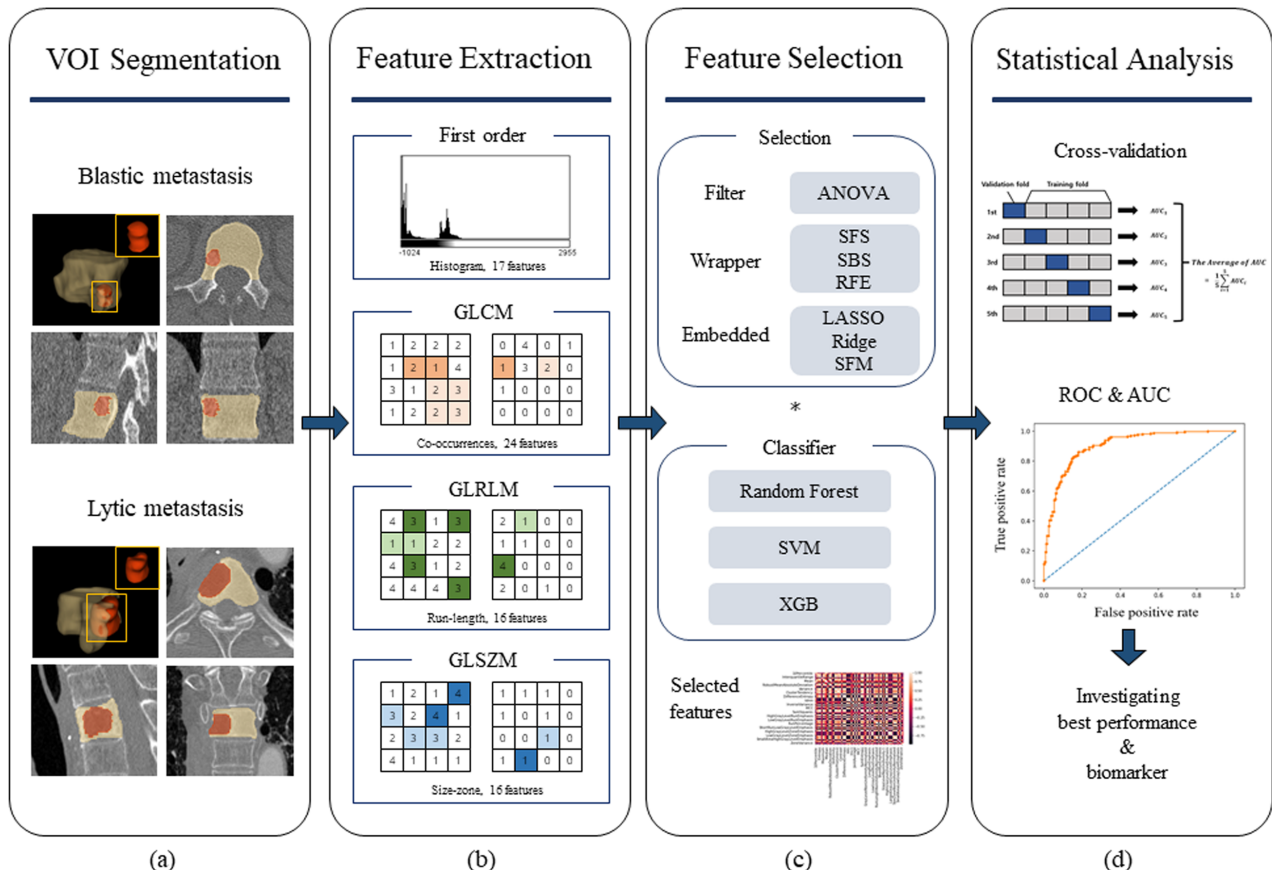
특징 선택은 예측 성능의 향상과 효율적인 머신러닝의 알고리즘 훈련을 위한 필수적인 단계이다. 특징선택 기법은 크게 필터방법(filtering), 래퍼방법(wrapper), 그리고 임베디드 방법(embedded)으로 크게 3가지로 분류된다. 필터방법은 다른 방법과 달리 기계학습 알고리즘과 독립적인 성향이 있으며, 본 연구에서 쓰인 일원분산분석(ANOVA, analysis of variance)은 두 개 이상 다수의 집단 간 분산을 분석하여 비교하는 방법이다. 분류 성능의 예측 정확도가 높은 것으로 알려진 래퍼방법의 sequential forward selection(SFS)은 머신러닝 모델에 입력되는 최적의 특징 조합을 찾기 위해 모델의 성능이 최대로 높아지는 순간까지 반복하여 특징을 입력하는 방법이다. 또한, sequential backward selection(SBS)은 모델에서 반복적으로 특징을 제거하여 최적의 특징 조합을 선택함으로써 모

텔의 성능을 높이는 방법이다. 그 외, 과적합을 줄이기 위해 내부적으로 패널티 성향을 지닌 임베디드 방법의 least absolute shrinkage and selection operator(LASSO)와 ridge regularization는 각각 계수 크기의 절대값과 제곱에 해당하는 패널티를 부여함으로써 정규화를 수행하는 방법이다. 본 연구에서는 1개의 필터방법의 ANOVA [20], 3개의 레퍼방법의 SFS, SBS[21], recursive feature elimination(RFE) with permutation feature importance [22], 3개의 임베디드방법의 LASSO, ridges regularization [23], select from model(SFM) with permutation feature importance [24]으로 총 7개의 특징선택 기법을 사용했다. 골 전이암 분류 성능을 비교함으로써 병변의 정확한 진단을 위한 생체지표를 평가하는 것을 목표로 두어 분류와 회귀가 가능한 extreme gradient boost(XGB)와 서로 다른 도메인(domain)에서 반복적으로 좋은 예측 성능을 나타내는 분류기인 random forest(RF), support vector machine(SVM)의 총 3개의 분류기를 사용하였다[25](그림 2c).

4. 통계 분석 방법

제시하는 연구에서는 머신러닝의 골 전이암 분류 성능을 평가하기 위해 의료진의 판독 결과(gold-standard)와 머신러닝의 예측결과를 비교하였다. 비교를 통해 얻어진 true positive(TP), false negative(FN), true negative(TN), false positive(FP)를 이용하여 정확도, 민감도(sensitivity), 특이도(specificity)를 얻었다. 그림 2d에서 제시한 바와 같이, 골 전이암 분류 성능의 평가를 위해 receiver operating characteristic (ROC) 분석을 수행하였으며, ROC 곡선의 면적을 표현하는 값인 area under the curve(AUC)를 통해 두 가지의 골 전이암 추체와 정상추체 분류 모델의 성능을 평가하였다. ROC분석에서 두 가지의 병변 및 정상을 분류할 수 있는 최적의 기준점(cut-off point)을 정하기 위해 두 가지의 병변과 정상은 각각 다른 값의 Youden index J(Youden's J statistic)가 선택되었다[26].

74



약어: VOI, volume of interest; GLCM, gray level co-occurrence matrix; GLRLM, gray level run length matrix; GLSZM, gray level size-zone matrix; ANOVA, Analysis of Variance; SFS, Sequential Forward Selection; SBS, Sequential Backward Selection; LASSO, Least Absolute Shrinkage and Selection Operator; ROC, receiver operating characteristic; AUC, area under the curve

그림 2. 제안하는 연구의 전체적인 흐름도

Fig. 2. The flowchart illustration of the proposed study

III. 연구 결과

1. 특징 선택 및 분류 접근법에 따른 골 전이암 분류 성능

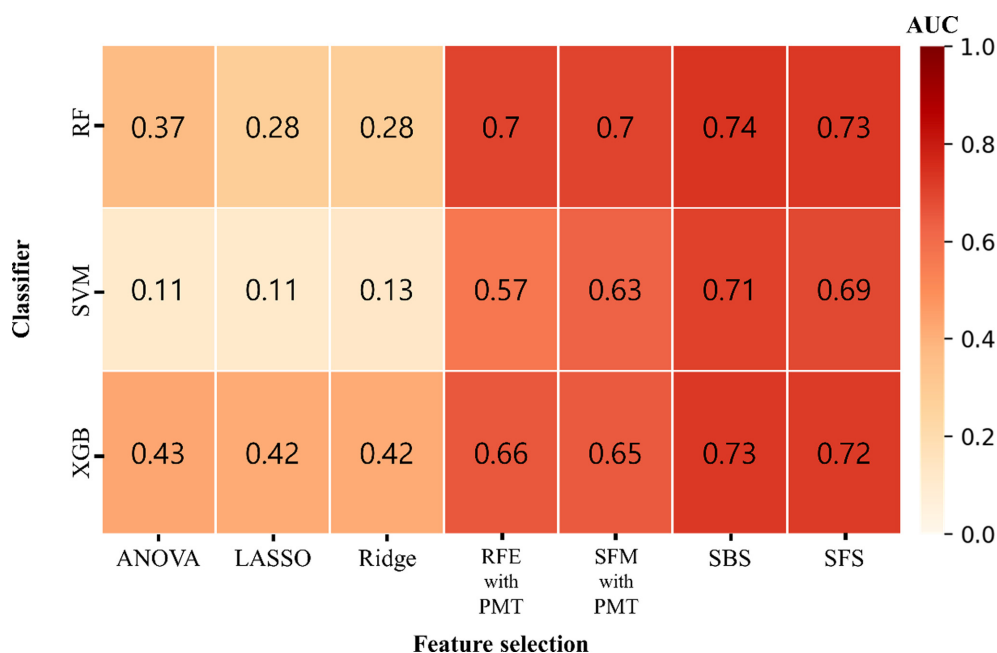
본 연구에서는 7가지 특징선택 기법과 3가지 분류기의 조합으로 얻어진 총 21가지 방법의 학습 결과와 전문의의 진단 결과를 비교 분석하여 두 가지의 골 전이암과 정상 추체의 분류 성능을 평가하였다(그림 2). 실험에 사용된 데이터는 세 가지의 종류로 구분이 되므로, 하나의 조합에서 얻어진 세 개의 이진-분류 결과로부터 얻어진 AUC 값의 평균 (micro-average ROC & AUC)을 분류기의 최종 분류 성능평가 기준으로 사용하였다.

그림 3에서 7가지의 특징 선택 기법과 3가지 분류기의 총 21가지의 조합에 대한 골 전이암 분류 성능 결과를 제시하였다. 분류성능의 비교 결과에 따르면, 특징선택기법 중 레퍼방법인 SBS와 RF의 분류기의 조합이 AUC 0.74로 우수한 성능까지 도달하였다. 특징선택기법들 중 레퍼방법과의 조합의 성능이 평균적으로 높은 AUC까지 도달하였으나, ANOVA, LASSO, 그리고 SVM의 조합은 AUC 0.11로 레퍼방법과 비교하였을 때 상대적으로 낮은 성능까지 도달하였다.

2. RF & SBS를 이용한 3진-분류 결과의 통계 분석

그림 4에서 골 전이암 분류에서 높은 성능을 보인 RF 및 SBS기법에서 골 모세포성 골전이를 포함하는 추체, 골 용해성 골전이를 포함하는 추체, 그리고 정상 추체의 분류 성능 평가 결과를 제시하였다. 제시한 결과는 3진-분류(three-class-classification) 결과에서 병변에 따른 이진분류 성능을 ROC 곡선을 이용하여 분석하였으며, 세 가지의 ROC 곡선의 평균 성능을 분석하기 위해 micro-average ROC 곡선 또한 제시하였다. 3진-분류 결과, 세 가지의 서로 다른 ROC 곡선으로부터 얻어진 AUC는 골 모세포성 골전이를 포함하는 추체, 골 용해성 골전이를 포함하는 추체, 정상 추체에서 각각 AUC 0.78, 0.65, 0.77까지 도달하였으며, micro-average ROC의 AUC는 0.74까지 도달하였다.

본 연구는 골 전이암 분류 성능의 향상과 학습 조합에 대한 신뢰도 평가를 위해 격자탐색(grid search)과 5겹 교차 검증(cross-validation)을 동시에 수행하였다(그림 2d). 표 1에서 RF와 SBS의 조합의 교차 검증에 대한 결과를 제시하였으며, 해당 조합에서 두 가지의 골 전이암을 가진 추체와 정상 추체의 분류성능을 평가하기 위해 5겹 교차 검증을 통한 5개의 데이터세트(dataset)에서의 예측결과의 성능을 평

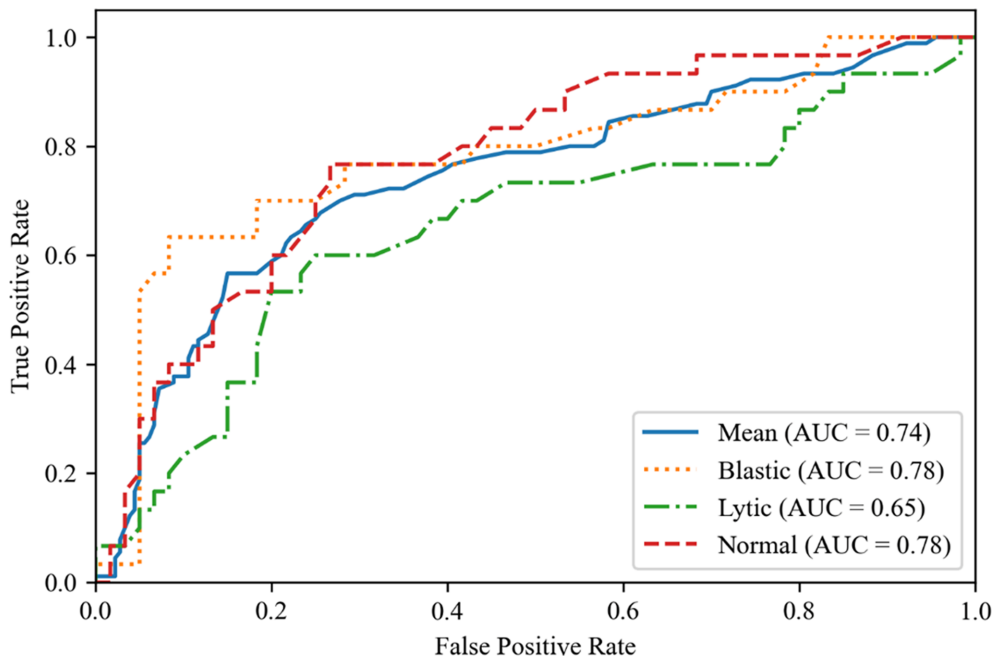


75

약어: XGB, extreme gradient boosting; RF, random forest; SVM, support vector machine; ANOVA, analysis of variance; LASSO, least absolute shrinkage and selection operator; RFE, recursive feature elimination; PMT, permutation feature importance; SFM, select from model; SBS, sequential backward selection; SFS, sequential forward selection; AUC, area under the curve

그림 3. 최적의 척추 전이 분류 성능을 확인하기 위한 세 가지의 분류기와 일곱 가지의 특징 선택 기법의 성능평가 분석 결과이며, 제시된 수치는 2가지의 골 전이암을 가진 추체와 정상 추체의 분류 성능의 micro-average AUC이다.

Fig. 3. The results of the ablation study using three-different classifiers and seven-different feature selection approaches for identifying the optimal spinal metastases classification performance. Presented numerical values indicate micro-average AUC score.



약어: ROC, receiver operation characteristic; SBS, sequential backward selection; XGB, extreme gradient boosting; RF, random forest

그림 4. 척추 전이 분류를 위한 RF 분류기와 SBS 특징선택 방법의 ROC 분석 결과. 골 용해성 골전이, 골 모세포성 골전이와 정상의 추체의 이진 분류 성능을 통해 각각 노란색, 녹색 및 빨간색 선으로 ROC 곡선을 제시하였다. 제시된 파란 선은 micro-average ROC 곡선이다.
Fig. 4. The results of ROC analysis that obtained via RF classifier and SBS feature selection for spinal metastases classification. The binary classification performances of blastic, lytic, and normal-healthy were indicated as yellow, green, and red-line, respectively. Additionally, the presented blue-line indicates micro-average ROC curve of three-different ROC curves of blastic, lytic, and normal-healthy

표 1. RF와 SBS 조합의 민감도, 특이도, 정확도, Youden index, AUC, 95% 신뢰구간의 분석결과. 가장 높은 값은 굵은 글자로 표시하였다.
Table 1. The sensitivity, specificity, accuracy, Youden index, AUC, and 95% CI of prediction results obtained via random forest classifier and sequential backward selection approaches. The highest values were indicated in bold.

	Sensitivity	Specificity	Accuracy	Youden index J	AUC	95% CI
Normal	0.77	0.73	0.74	0.50	0.77±0.12	0.68-0.86
Lytic	0.60	0.75	0.63	0.35	0.65±0.12	0.54-0.75
Blastic	0.63	0.92	0.80	0.55	0.78±0.09	0.68-0.86

약어: AUC, area under the curve; CI, confidence interval

가하였다. 분류 성능 평가결과, 평균 민감도, 평균 특이도, 평균 정확도, Youden index, 그리고 95% 신뢰구간(confidence interval, CI) 값을 얻었다. 골 모세포성 골 전이암을 포함하는 추체의 경우 민감도, 특이도, 정확도, Youden index 값은 각각 0.63, 0.92, 0.80, 0.55, 골 용해성 골전이를 포함하는 추체는 0.60, 0.75, 0.63, 0.35, 정상 추체는 0.77, 0.73, 0.74, 0.50 까지 도달하였다.

그림 5는 RF와 SBS 조합에서 5점 교차검증한 모델의 평균적 모델에서 특징의 중요도 값을 나타낸 그래프이다. 머신러닝에 사용되는 특징이 많을수록 모델의 복잡성과 오류 발생의 가능성이 높아 모델의 성능이 낮게 평가될 수 있다. 추출한 특징의 개수가 많을 경우, 특징의 중요도 분석을 통해

유의미한 특징들을 선정하는 것이 중요하다. 골 전이암 분류 성능에 유의한 영향을 미치는 특징의 중요도의 값을 구해 중요도가 높은 특징을 골 전이암 추체를 분류하기 위한 생체 지표로 선정하였다. 골 전이암 분류 성능에 중요도가 높은 특징들은 energy, entropy, kurtosis, variance, long run low gray level emphasis, gray level variance, large area low gray level emphasis, idn, 90 percentile, sumsquares이다. 그 중, 상위 3가지의 특징들은 90 percentile, kurtosis, energy이며, 해당 특징들은 각각 관심영역 내부에서 복셀-강도의 90-백분위 값, 히스토그램의 첨도 값, 3차원 관심영역 내부 복셀-강도 값의 총합을 의미한다.

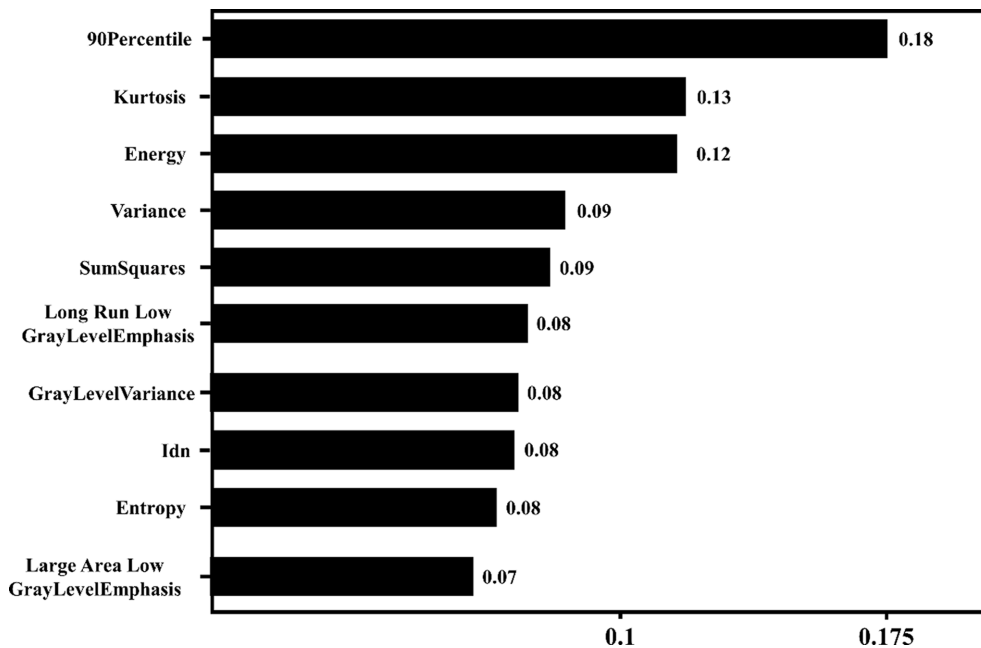


그림 5. Sequential backward selection 특징선택 기법을 이용한 RF 분류기의 특징 중요도 분석 결과

Fig. 5. The feature importance analysis of RF classifier using sequential backward selection approach

IV. 고찰 및 결론

본 논문에서는 골 전이암 추체를 정량적으로 분류를 위해 CT영상에서 3차원-라디오믹스 기반의 특징을 추출하여 7가지의 특징 선택기법과 3가지의 분류기를 이용하여 기계학습 결과 분석을 수행하였다. 그 뿐만 아니라, 골 전이암 분류에 영향을 크게 미치는 특징의 중요도 값을 조사하여 3번 째로 가장 높은 중요도를 가진 특징까지를 골 전이암 분류의 생체지표로 선정하였다. 선택된 특징의 생체지표로서의 사용 가능성을 평가하기 위해 총 21가지 조합의 머신러닝 예측 결과와 의료진의 판독결과를 비교 분석하여 골 전이암 분류 성능을 평가하였다. 분석결과, RF와 레퍼방법이 비교적 다른 조합들보다 높은 성능을 보였으며, 분류기에서 RF, 특징 선택 기법에서 SBS 조합에서 0.74의 AUC까지 도달하였다. RF와 SBS 조합의 골 전이암 분류 성능 분석결과, 골 모세포성 골전이 추체의 분류 성능평가에서 골 전이암 분류를 위한 정확도를 높이기 위해 높은 특이성에 비해 낮은 민감성을 이용하였으며, 특이도와 민감도의 값은 각각 0.92, 0.63이다. 해당 조합에서 골 전이암 분류에 유의미한 특징은 90 percentile, kurtosis, energy 값을 확인할 수 있었다.

골 전이암 분류 성능이 우수한 SBS와 RF의 조합과 반대로 성능이 낮은 조합은 ANOVA 특징선택 방법과 SVM 분류기의 조합이었다. ANOVA 분석은 2가지 골 전이암을 가진 추체와 건강한 정상 추체에서 분산분석을 수행하였을 때, 세 가지 중 최소 한 가지 이상의 군에서 분산의 차이를

가질 경우, 유의한 p 값(p -value)이 얻어진다[27]. 제시한 연구에서 ANOVA 분석 방법을 이용한 필터 특징선택 방법에서는 유의한 p 값을 $p < 0.05$ 로 설정하였다. ANOVA 분석 결과, 실험에서 사용된 데이터에서 3진 분류했을 때, 골 용해성 골전이 추체와 정상 추체에서는 유의한 분산의 차이가 없었으나($p=0.236$), 골 모세포성 골전이 추체가 정상 및 골 용해성 골전이 추체와 유일하게 유의한 차이를 보였다. 그러나, 성공적인 3진 분류 작업을 위해서는 골 모세포성 골전이, 골 용해성 골전이, 정상 추체에서 모두 유의한 차이를 보이는 특징을 선택해야 하므로 ANOVA이 3가지 추체를 모두 분류하기에 적합한 특징을 선택하기 어려웠을 것으로 판단된다. 마찬가지로, l1 및 l2 정규화 방법을 이용하는 LASSO와 ridge 특징선택 방법도 모든 분류기와의 조합에서 낮은 성능까지 도달하였다. LASSO 기법의 경우, 특징들 간에 상관관계가 클 경우 유의한 차이를 보이는 특징만을 선택함과 동시에 나머지 특징 값을 0으로 정규화 하므로, 골 전이암 분류에 어느 정도 필요한 특징 정보의 손실로 인해 분류 정확도가 낮은 것으로 판단된다. Ridge 기법의 경우, 모든 특징을 유지함과 동시에 유의하지 않은 특징 값을 점진적으로 감소시킨다. 이에 따라, 유의하지 않은 정보를 모두 보존하기 때문에 낮은 분류 성능을 도달한 것으로 판단된다.

반면, 세 가지의 레퍼 특징선택 방법들은 다른 특징선택 방법들과 비교하여 높은 성능까지 도달함을 증명하였다. 제시한 연구에서는 레퍼방법의 최적화 과정에서 특징을 추가하거나 빼는 서브세트(subset)를 1개로 설정하였다. 본 방

법은 분류기의 성능이 최적화되기까지 1개의 특징 서브세트를 반복적으로 추가하거나 제거하는 방법으로, 제시한 연구에서는 총 73번의 반복 처리를 통하여 모델을 최적화하였다. 본 연구에서는 총 73개인 작은 수의 라디오믹스 특징을 이용하였기 때문에, 결정 경계의 최적화 작업에 비교적 시간 소모가 존재함에도 불구하고, 높은 분류성능까지 도달하는 레퍼방법이 적합했던 것으로 판단된다.

제시한 실험결과에서는 SVM이 모든 특징선택 기법 및 분류기에서 가장 낮은 골 전이암 분류성능까지 도달하였다. 실험에 사용된 SVM에서는 선형 커널(linear kernel)을 이용하여 결정 경계(decision boundary)를 최적화하였다. 반면, XGB와 RF는 비-선형(non-linear)적인 결정 경계를 최적화하는 기법이다. 실험에 사용된 데이터 및 실험결과를 분석하였을 때, 다음과 같은 이유로 선형 결정 경계의 최적화가 어려웠던 것으로 판단된다: 1) 제안하는 연구에서는 세 가지의 그룹을 동시에 분류하는 3진-분류 작업을 수행하는 것이 목표이므로, 결정 경계가 선형일 경우, 복잡한 특징 값의 분포를 구분하는 작업에는 적합하지 않다. 2) ANOVA 분석 결과에서 밝혀낸 바와 같이, 실험에서 사용된 blastic 추체만이 나머지 두 가지 군집과 유의한 분산의 차이를 보였다. 그러나, 3진-분류 작업을 위해서는 blastic뿐만 아니라, 나머지 두 군 또한 분류하는 결정 경계가 필요하기 때문에 비-선형 결정 경계가 본 실험에서 사용된 데이터에 적합하다. 또한 연구 결과 RF와 XGB의 성능이 대략 비슷하게 나타났지만 RF의 성능이 조금 더 좋게 나온 것을 감안하면, XGB보다는 RF와의 조합이 적정적합(Generalized-fitted) 일어났음을 짐작할 수 있다.

분류기 및 특징선택을 통해 골 전이암 분류에 기여도가 높은 특징이 90-percentile, kurtosis, energy임을 증명하였다. 기존 임상에서의 골 전이암 진단은 CT영상에서 의사의 육안을 기반으로 골전이 병변 조직의 기질(matrix)을 가늠하거나 골 파괴의 정도를 파악하여 골 전이암의 위험성을 파악하였다[28]. 기존의 척추 골 전이암 임상-의사결정(clinical decision)에 쓰인 병리학적 진단방법에 따르면, 추체 내부의 강도 값 및 골밀도가 골전이 진단에 유의한 특징으로 알려진 바 있다. 따라서, 실험을 통해 선택된 90-percentile, kurtosis, energy 특징들이 골전이 진단에 사용 가능성이 높은 것으로 판단된다.

본 연구에서는 골 전이암 분류 성능에 대한 신뢰도 높은 평가를 위해 5점-교차검증을 하여 분류 성능 평가의 편중(bias)을 방지하고 골 전이암 분류 정확도를 향상시켰다. RF 분류기와 SBS 특징선택 기법에서 쓰인 중요도 높은 특징을 골 전이암 여부를 정량적으로 측정 및 평가할 수 있는 생체 지표로 선정하여 골 전이암 조기 진단의 가능성을 조사하였다. 제안한 특징을 이용하여 향후 CT영상에서 척추 전이성

암종의 조기진단 및 정량적인 추적 관찰에 큰 도움이 될 것으로 기대한다.

References

- [1] Mastro AM, Gay CV, Welch DR. The skeleton as a unique environment for breast cancer cells. *Clinical & Experimental Metastasis*. 2003;20(3):275-84.
- [2] David Roodman G, Silbermann R. Mechanisms of osteolytic and osteoblastic skeletal lesions. *Bonekey Rep*. 2015;4:753.
- [3] Shin SO, Kim SK, Kim MS. Pallative effect of radiation therapy in management of symptomatic osseous metastases. *Yeungnam University Journal of Medicine*. 1992;9(1):102.
- [4] Sohn SK et al. Collective Review of Cases of Spinal Metastases. *The Korean Orthopaedic*. 1988;23(4):1087-96.
- [5] Batson OB, MA. The Fuction of the Vertebral Veins and Their Role in the Spread of Metastases. *Annals of Sugery*. 1940;112(1):138-49.
- [6] Liaw C-C, et al. Hepatocellular Carcinoma Presenting as Bone Metastasis. *Cancer*. 1989; 64(8):1753-57.
- [7] Park J-M. Interventional treatments for cancer pain due to bone metastasis. *Anesthesia and Pain Medicine*. 2015;10(3):149-64.
- [8] Koom WS, et al. Radiation Therapy for Bone Metastasis from Hepatocellular Carcinoma. *Clinical and Molecular Hepatology*. 2002;8(3):304-11.
- [9] Heindel W. et al. The diagnostic imaging of bone metastases. *Deutsches Arzteblatt international*. 2014;111(44):741-47.
- [10] HM, B.-C., J.-Z. O, M. JD. Diagnosis and Treatment Options of Spinal Metastases. *Rev Invest Clin*. 2015;67(3):140-57.
- [11] Wang Z, et al. Exploiting machine learning for predicting skeletal-related events in cancer patients with bone metastases. *Oncotarget*. 2015;11(7):12612-22.
- [12] Eun PJ, KH Sung Radiomics as a Quantitative Imaging Biomarker: Practical Considerations and the Current Standpoint in Neuro-oncologic Studies. *Nuclear medicine and molecular imaging : NMML*. 2018;52(2):99-108.
- [13] Ahn SJ, et al. Contrast-enhanced T1-weighted image radiomics of brain metastases may predict EGFR mutation status in primary lung cancer. *Scientific Reports*. 2020;10(1):8905.
- [14] Kim KH. Non-small cell lung cancer recurrence prediction model using deep learning-based radiomics. *Conf Proc 대한 기계학회 춘추학술대회*, 2020. p. 53.
- [15] Lee G, Bak SH, Lee HY. CT Radiomics in Thoracic Oncology: Technique and Clinical Applications. *Nuclear Medicine and Molecular Imaging*. 2018;52(2):91-8.
- [16] A F., et al. 3D Slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magn Reson Imaging*. 2012; 30(9):1323-41.
- [17] Yang Y, et al. Optimizing Texture Retrieving Model for Multimodal MR Image-Based Support Vector Machine for Classifying Glioma. *J Magn Reson Imaging*. 2019;49(5):1263-74.
- [18] McKay C, Fujinaga I, Depalle, P. jAudio: A feature extraction library. in *Proceedings of the International Conference on Music Information Retrieval*. 2005.
- [19] Drotár P, Gazda J, Smékal Z. An experimental comparison of feature selection methods on two-class biomedical datasets. *Computers in Biology and Medicine*. 2015;66:1-10.
- [20] Küffner R, et al. Inferring gene regulatory networks by ANOVA.

- Bioinformatics. 2012;28(10):1376-82.
- [21] Chandrashekar G, Sahin F. A survey on feature selection methods. *Computers & Electrical Engineering*. 2014;40(1):16-28.
- [22] Yan K, Zhang D. Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical*. 2015;212:353-63.
- [23] Lu M. Embedded feature selection accounting for unknown data heterogeneity. *Expert Systems with Applications*. 2019;119:350-61.
- [24] Altmann A., et al. Permutation importance: a corrected feature importance measure. *Bioinformatics*. 2010;26(10):1340-47.
- [25] Verma C, Illés Z, Stofová V. Real-time classification of national and international students for ICT and mobile technology: an experimental study on Indian and Hungarian University. *Journal of Physics: Conference Series*. 2020;1432:12091.
- [26] Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biom J*. 2005;47(4):458-72.
- [27] Armstrong RA, Slade SV, Eperjesi F. An introduction to analysis of variance (ANOVA) with special reference to data from clinical experiments in optometry. *Ophthalmic and Physiological Optics*. 2000;20(3):235-41.
- [28] Shin DS, Ryu SM, Park CH. The Diagnostic Strategy for Malignant Bone Tumors. *The Korean Orthopaedic*. 2015;50(6):429-37.