

텍스트 마이닝 기법을 활용한 어깨 재활 연구분야 동향과 키워드 모델링

김준희 · 정성훈¹ · 황의재[†]
코리아테크, ¹연세대학교 물리치료학과

The Research Trends and Keywords Modeling of Shoulder Rehabilitation using the Text-mining Technique

Jun-hee Kim, PT, PhD · Sung-hoon Jung, PT, PhD¹ · Ui-jae Hwang, PT, PhD[†]

Research & Development, KOREATECH Corporation,

¹Department of Physical Therapy, Yonsei University

Received: April 12, 2021 / Revised: April 14, 2021 / Accepted: May 02, 2021

© 2021 J Korean Soc Phys Med

| Abstract |

PURPOSE: This study analyzed the trends and characteristics of shoulder rehabilitation research through keyword analysis, and their relationships were modeled using text mining techniques.

METHODS: Abstract data of 10,121 articles in which abstracts were registered on the MEDLINE of PubMed with 'shoulder' and 'rehabilitation' as keywords were collected using python. By analyzing the frequency of words, 10 keywords were selected in the order of the highest frequency. Word-embedding was performed using the word2vec technique to analyze the similarity of words. In addition, the groups were classified and analyzed based on the distance (cosine similarity) through the t-SNE technique.

RESULTS: The number of studies related to shoulder rehabilitation is increasing year after year, keywords most frequently used in relation to shoulder rehabilitation studies are 'patient', 'pain', and 'treatment'. The word2vec results showed that the words were highly correlated with 12 keywords from studies related to shoulder rehabilitation. Furthermore, through t-SNE, the keywords of the studies were divided into 5 groups.

CONCLUSION: This study was the first study to model the keywords and their relationships that make up the abstracts of research in the MEDLINE of Pub Med related to 'shoulder' and 'rehabilitation' using text-mining techniques. The results of this study will help increase the diversifying research topics of shoulder rehabilitation studies to be conducted in the future.

Key Words: Research trend, Shoulder rehabilitation, t-SNE, Text-mining, word2vec

[†]Corresponding Author : Ui-jae Hwang

Hwangu33@nate.com, <https://orcid.org/0000-0002-2050-5503>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

I. 서론

어깨부위의 병변은 나이에 따라 약 9-25% 정도에서 관찰되며, 성인의 경우 많게는 66.7% 정도가 일생동안 어깨 통증을 한번 이상 경험하는 것으로 알려져 있다 [1]. 건강보험심사평가원에서 공개하고 있는 통계지표에 따르면, 어깨 부위의 병변은 2019년 40대 입원 환자가 가장 많았던 질병 가운데 5번째를 차지하였으며, 2010년에 발생한 어깨 병변 환자의 수는 약 200만명에서 2019년에는 그 수가 약 230만명으로 매년 꾸준히 증가하였다. 어깨 부위는 일상생활을 수행하는 동안 가장 빈번하게 사용하는 관절이기에 어깨 부위 병변은 스포츠 활동을 어렵게 할 뿐만 아니라, 직업 활동을 어렵게 하고 나아가 일상생활에서 동작을 수행하는 데 있어 어려움을 발생 시킬 수 있다[2-5].

어깨 부위에서 발생하는 병변으로는 주로 유착성 관절주머니염, 돌림근띠 파열, 어깨뼈 아래쪽 돌림 증후군, 근막통증 증후군, 어깨관절 충돌 증후군, 어깨관절 석회화 건염 등이 있다[6,7]. 이러한 어깨 부위에서 발생하는 병변들을 개선하거나 치료하기 위해 수술적 치료, 재활 운동, 물리적 인자 치료, 침술, 보조기 등과 같은 다양한 방법들이 연구되어왔다[7-11]. 특히, 산업 구조의 변화와 맞추어 직업 특성과 연관된 반복적인 작업, 무리한 힘의 사용, 부적합한 자세 등이 어깨 부위 병변을 증가시켜 왔고 여가 활동의 증가로 인한 스포츠 활동도 어깨 부위 병변의 발병을 증가시키고 있다 [12-17]. 이러한 사회적 흐름과 함께 어깨 재활에 관한 관심들은 점차 증가하였고 이에 대한 연구들이 꾸준히 증가하고 있다.

많은 연구자들은 문자 형태로 문서에 존재하는 다양한 정보에 대한 해석과 분석을 통하여 가치 있는 정보를 찾아내기 위해 노력을 하고 있다. 텍스트 마이닝은 통계적 패턴 학습, 자연어 처리 기법과 같은 방식을 기반으로 문자 데이터로부터 의미 있는 정보들을 얻어내는 작업이다[18]. 텍스트 마이닝 기법은 산업공학, 경제학, 건설 등 다양한 분야에서 사용되고 있으며, 이에 대한 방법들도 끊임없이 발전하고 있다[19,20]. 이러한 방법들 덕분에 기존의 방식으로 해결해 왔던 주제로부터 더욱 다양

하고 넓은 주제에 대한 분석이 가능해지고 있다.

기존 재활과 관련된 분석 연구들은 치료 방법들에 대한 선행 연구들을 기반으로 그 효과를 검증하는 연구들이 주를 이루었다. 하지만, 주로 대상자들의 특성, 치료 방법, 치료의 효과 데이터만을 사용하여 분석하였으며, 문자 데이터를 이용하여 분석에 사용한 경우는 거의 없었다. 텍스트 마이닝적 접근을 통해 포털사이트에서 검색되는 주요 단어들의 관계를 분석하여 ‘작업치료’라는 단어의 대중적 의미를 파악했던 연구와 한국언어치료학회지에 게재되어 있는 연구들의 주제 단어들을 모델링하여 그 트렌드를 분석한 재활 분야 연구도 존재하지만, 이들 연구들은 특정 학문 분야에 대한 대중 인식 파악 혹은 특정 학술지 내의 경향성을 분석했다 [21,22]. 그래서, 본 연구에서는 텍스트 마이닝적 접근을 통해 최대 의학 서지 데이터베이스인 MEDLINE에 등록되어 있는 어깨 재활 관련 연구들의 전반적인 동향과 특징들을 제시하고, 초록 데이터를 이용하여 어깨 재활에 관한 키워드와 이들의 관계를 모델링 하고자 하였다.

II. 연구방법

1. 자료 수집

어깨 재활 연구들을 대표할 수 있는 논문들을 선정하기 위해 펌메드 사이트 (<https://pubmed.ncbi.nlm.nih.gov/>)에서 검색어를 ‘어깨’와 ‘재활’ (“shoulder” and “rehabilitation”)로 하여 2021년 3월 7일 기준으로 초록 데이터가 등록되어 있는 10,338편의 논문들에 대한 제목, 저자, 저널명, 초록, 출판일 자료를 수집하였다. 논문들의 데이터 수집은 python의 selenium 모듈과 BeautifulSoup를 이용하여 작성한 프로그램으로 수행하였다. 이 중 중복 논문들을 제외한 10,121편의 논문들을 활용하여 키워드 분석과 워드-임베딩 분석을 수행하였다.

2. 명사 단어 전처리

키워드 분석과 워드-임베딩 분석에 있어 불필요하게 여겨지는 단어들에 대한 제거 작업을 수행하였다. 초록에 문단 제목으로 쓰이는 단어 (background, results,

conclusion 등), 연구 방법론적으로 쓰이는 단어 (study, methods, group, control 등)들이 해당되었다. 그리고, 복수형 및 단수형 단어들의 통합과 같은 추가 작업들을 수행하였다. 그리고, 워드-임베딩을 수행하기 전 데이터를 단어별로 분류하는 작업이 필요하다. 특히, 수집된 데이터 중에서 분석을 하는 데 있어 중요하지 않은 단어들이 관사 (a, an, the), 전치사 (of, on, for), 대명사 (it, he, her) 등을 제거해야 하고 분류해야 한다. 이와 함께, re 모듈을 활용하여 분석에 활용되지 않는 특수문자와 숫자 데이터들을 제거하였다.

3. 키워드 분석

Python의 pandas 모듈과 카운터 함수를 이용하여 초록에 등장하는 단어들의 빈도를 분석하고 빈도가 높은 순서대로 단어를 정렬해주는 프로그램을 작성하였다. 분석결과와 정확성 향상을 위해 동일한 의미를 가진 단어들은 같은 단어로 병합하였다. 이 때, 빈도수가 높은 단어들 중에서 검색어인 ‘어깨’와 ‘재활’을 제외한 10개의 단어를 키워드로 분석하였다.

4. 워드-임베딩

워드-임베딩은 단어를 벡터의 형태로 순서와 의미를 표현하는 기법으로 머신러닝 알고리즘에 의해 단어들의 의미 관계가 보존되도록 하는 방법이다. 이러한 방법으로는 word2vec 모델이 가장 대표적인 방법으로 Mikolov 외 4인이 제안하였다[23]. word2vec는 특정 embedding 공간상에서 같은 맥락 (context)을 갖는 단어들은 가까운 거리를 가진다는 전제로 시작한다. 이를 통해, word2vec 표현법으로 주어진 문장에 대한 문법적 해석과 단어들의 거리를 통한 의미론적 추론이 가능해진다. 즉, 머신 러닝 알고리즘에 의해 의미론적 관계가 보존되도록 벡터화 된다. 예를 들어, ‘한국’, ‘서울’, ‘일본’, ‘도쿄’에 해당하는 벡터 존재한다고 할 때, ‘한국’ - ‘서울’ + ‘일본’ = ‘도쿄’의 관계가 보존되도록 하는 것이다.

5. Word2vec

word2vec에서는 데이터를 학습하고 예측하는 방법

은 CBOW (Continuous Bag-of-words) 와 skip-gram이 있다. CBOW 방법은 2k개의 주변 단어가 존재할 때 그 중심에 특정 단어가 나타날 조건부 확률을 계산하는 것이다. 반면, skip-gram 방법은 중심 단어가 주어졌을 때 특정 조합을 가진 주변 단어 2k개가 나타날 조건부 확률을 계산하여 워드-임베딩을 수행한다. 두 가지 모형 가운데 skip-gram 방법이 워드-임베딩 시 속도는 느리지만 정확도가 높다고 보고되었다[24]. 본 연구는 python의 gensim 모듈의 word2vec 함수를 이용하여 Skip-Gram 방식에 단어 벡터 크기 300차원, 문맥 윈도우 10으로 설정하여 진행하였다[25]. 이를 통해 단어들을 R^{300} 공간상에 300차원 벡터로 학습하였다. 이와 함께, `wv.most_similar` 함수를 이용하여 키워드들과 연관성이 가장 높은 단어들을 코사인 유사도 (Cosine similarity)로 계산하였다. 코사인 유사도 계산은 벡터로 이루어진 두 단어 사이의 코사인 각도를 구하고 이를 통해 단어들이 얼마나 유사한지 계산하는 방법이다. 즉, 이 측정 방법은 두 벡터가 방향이 같을수록 유사하다고 판단한다. 이때 계산한 값이 1에 가까울수록 두 단어가 양의 유사도를 갖고, 0에 가까울 때는 유사도가 없으며, -1에 가까울수록 음의 유사도를 지닌다.

5. t-SNE

word2vec을 통하여 다차원으로 구성된 자료들을 시각화하기 위해서는 적은 차원으로 축소해야 한다. PCA는 기존에 많이 사용되던 선형적인 차원축소 방법이라면, 최근에는 t-SNE처럼 머신 러닝에 의한 비선형적 차원 축소 방법들도 많이 개발되어 사용되고 있다. 특히, Maaten과 Hilton[26]이 제안한 t-SNE는 데이터상의 군집들을 다른 방법들보다 더 잘 보여주는 경향이 있다고 보고되었다. 본 연구에서는 t-SNE를 이용하여 300차원 공간의 단어들을 2차원 공간으로 시각화 하였다. 그 다음 ‘어깨’, ‘재활’ 및 키워드 10개 단어들과 데이터 간 거리가 가까운 단어들을 분석하였다. 이때 데이터 간 거리를 나타내는 값이 작을수록 데이터 간의 연관성이 높은 것으로 판단한다.

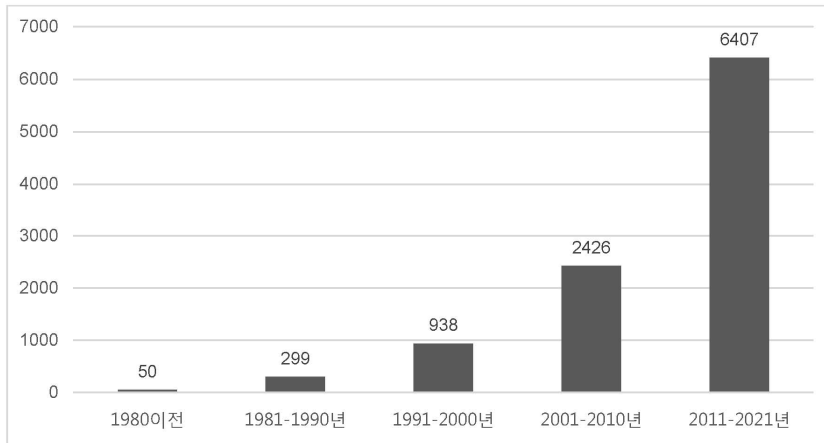


Fig. 1. Number of studies in 10 years.

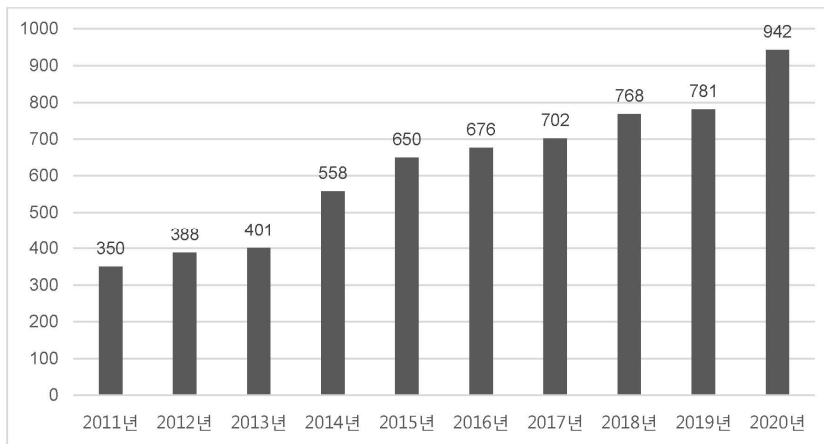


Fig. 2. Number of studies per year within the last 10 years.

III. 연구결과

1. 자료수집 결과

프로그램 실행 결과 논문은 총 10,338편이 검색되었으며 이 가운데 중복을 제외한 조사대상 논문 건수는 1971년부터 2021년 3월 7일까지 총 10,121편이 조사되었다.

Fig. 1과 같이 10년 단위로 구분된 논문들의 숫자를 살펴보면 어깨 재활 관련 연구는 1981년부터 1990년까지 출판된 299편의 연구들이 데이터베이스에 등록되었으나 이후 급격히 증가하기 시작하여 1991년에서 2000

년까지 출판된 938편의 연구가 데이터베이스에 추가되었다. 2001년부터 2010년까지 출판된 연구들은 이전보다 훨씬 증가한 2426편의 연구가 데이터베이스에 등록되었고, 2011년부터 현재까지 출판된 6407편의 연구가 등록되어 이와 관련된 연구들이 급격히 증가하고 있다는 사실을 유추해볼 수 있다. 특히, 2010년대에 이르러서는 Fig. 2와 같이 매년 발표되는 ‘어깨’ (shoulder), ‘재활’ (rehabilitation) 관련 연구들의 숫자가 2011년 350편을 시작으로 2020년 942편에 이르기까지 해마다 증가하는 것을 볼 수 있다.

Table 1. 10 Top Keywords and Frequency

-1980		1981-1990		1991-2000		2001-2010		2011-2021		Total	
Keyword	Count	Keyword	Count	Keyword	Count	Keyword	Count	Keyword	Count	Keyword	Count
patient	121	patient	544	patient	2,061	patient	5,560	patient	15,516	patient	23,802
treatment	29	muscle	220	pain	859	pain	3,306	pain	10,339	pain	14,653
head	28	treatment	172	muscle	704	muscle	2,138	treatment	5,305	treatment	8,206
muscle	24	joint	155	treatment	677	treatment	2,023	score	5,175	muscle	7,957
arm	23	injury	146	injury	634	score	1,519	muscle	4,871	score	6,918
motion	21	pain	132	joint	536	joint	1,461	outcome	4,810	outcome	6,395
joint	20	activity	123	motion	461	degree	1,377	exercise	3,904	injury	5,665
function	18	motion	119	activity	389	motion	1,328	upper	3,724	exercise	5,598
pain	17	cuff	109	degree	386	exercise	1,280	injury	3,691	motion	5,542
elbow	17	rotator	104	cuff	379	range	1,274	motion	3,613	upper	5,384

2. 키워드 분석 결과

Table 1은 분석대상이 된 10,121편의 논문의 초록 텍스트를 이용하여 상위 10개의 키워드를 10년 단위로 분석하였다. 현재까지 발표된 ‘어깨’ (shoulder), ‘재활’ (rehabilitation) 관련 해외 연구들의 초록에서 가장 높은 빈도를 나타낸 단어는 ‘환자’ (patient), ‘통증’ (pain), ‘치료’ (treatment)인 것으로 나타났다. 그 다음으로는 ‘움직임’ (range, motion), ‘관절’ (joint), ‘근육’ (muscle)과 같은 단어들이 높은 빈도수를 나타냈다. 그리고 1981년도부터는 ‘활동’ (activity)이라는 단어가 키워드에 포함되었다. 그리고, 2001년부터는 ‘점수’ (score)라는 단어가 키워드에 포함되어 나타났다. 2001년부터 ‘운동’ (exercise) 단어가 점차적으로 증가하기 시작하였다. 2011년도부터 시작된 연구에서는 ‘결과’ (outcome)라는 단어의 빈도가 증가하였다.

3. 워드-임베딩

총 10,121편의 논문에 존재하는 초록 단어들을 워드-임베딩을 통해 300차원 공간으로 관계성이 보존되도록 표현하였다. 다음 Table 2는 가장 높은 빈도수를 나타낸 것으로 분석된 10개의 키워드들과 ‘어깨’ (shoulder), ‘재활’ (rehabilitation)과 가장 연관성 높은 단어들을 1위부터 10위까지 나타냈다. 각 키워드들과 관련된 단어들을 통하여 이들이 어떠한 개념과 관련성을 가지고 연구되

었는지 확인할 수 있다.

‘어깨’ (shoulder)는 ‘팔꿈치’ (elbow), ‘오목위팔’ (glenohumeral), ‘팔’ (arm) 등의 순서로 상지의 다른 부위를 나타내는 단어들이 높은 유사도를 나타냈다. ‘재활’ (rehabilitation)은 ‘프로그램’ (program), ‘치료’ (therapy), ‘관리’ (management) 등의 단어들과 높은 유사도를 나타냈다. ‘환자’ (patient)의 경우는 ‘대상자’ (subject), ‘케이스’ (case) 등의 단어들과 높은 유사도를 나타냈다. ‘통증’ (pain)은 ‘증상’ (symptom), ‘장애’ (disability), ‘호소 증상’ (complaint)과 같은 단어들이 높은 유사도를 나타냈다. ‘근육’ (muscle)의 경우 ‘세모근’ (trapezius), ‘근전도’ (emg), ‘삼각근’ (deltoid), ‘가시아래근’ (infraspinatus), ‘활성도’ (activation), ‘앞톱니근’ (serratus) 등과 같은 단어들이 높은 유사도를 나타냈다. ‘점수’ (score)는 평가 도구인 ‘시각통증척도’ (Visual Analogue Scale; VAS), ‘미국 어깨 및 팔꿈치 외과 의사 점수’ (American Shoulder and Elbow Surgeons Score; ASES), ‘상지 기능장애 평가도구’ (Disabilities of the arm, shoulder, and hand; DASH) 등이 높은 유사도를 나타냈고, ‘결과’ (outcome)는 ‘측정’ (measure), ‘기능적’ (functional), ‘점수’ (score), ‘장애’ (disability) 단어들과 높은 유사도를 나타냈다. ‘손상’ (injury)은 ‘외상’ (trauma), ‘운동선수’ (athlete), ‘스포츠’ (sport), ‘병변’ (lesion) 등이 높은 유사도를 나타냈다. ‘운동’ (exercise)은 ‘근력 강화’ (strengthening), ‘프로그

Table 2. 10 Word2vec Results of 12 Keywords

Rank	shoulder	rehabilitation	patient	pain	treatment	muscle	score	outcome	injury	exercise	motion	upper
1	elbow	program	subject	symptom	management	trapezius	scale	measure	trauma	strengthening	rom	extremity
	.404	.525	.436	.473	.564	.528	.556	.419	.453	.563	.539	.594
2	glenohumeral	therapy	case	disability	intervention	emg	ases	functional	athlete	program	range	limb
	.399	.472	.411	.460	.523	.467	.548	.355	.420	.471	.507	.562
3	pain	management	treated	complaint	therapy	deltoid	dash	score	sport	stretching	movement	motor
	.389	.429	.409	.425	.483	.458	.531	.347	.378	.449	.478	.400
4	joint	treatment	year	intensity	conservative	infraspinatus	vas	disability	lesion	resistance	rotation	muscle
	.366	.414	.398	.400	.433	.453	.527	.337	.365	.437	.430	.385
5	arm	care	enrolled	shoulder	rehabilitation	activation	constant	follow	overuse	training	passive	lower
	.351	.402	.398	.394	.413	.452	.507	.331	.361	.412	.422	.357
6	motion	physiotherapy	participant	problem	treated	serratus	murley	murley	dislocation	supervised	abduction	hand
	.351	.398	.390	.364	.379	.448	.492	.327	.360	.412	.396	.344
7	range	programme	month	chronic	nonoperative	musculature	ucla	status	league	programme	elevation	arm
	.323	.391	.383	.338	.378	.419	.481	.326	.348	.407	.392	.334
8	wrist	exercise	thirty	painful	patient	electromyographic	sf	primary	player	rehabilitation	flexion	ue
	.315	.373	.378	.336	.365	.395	.447	.320	.344	.370	.339	.311
9	hip	strengthening	twenty	headache	operative	contraction	p	vas	sprain	therapy	shoulder	function
	.312	.372	.377	.331	.363	.387	.447	.315	.340	.367	.338	.299
10	rotation	protocol	forty	neck	treat	upper	quickdash	clinical	competition	home	external	muscular
	.302	.367	.375	.327	.360	.387	.436	.313	.338	.365	.330	.295

램’ (program), ‘신장 운동’ (stretching), ‘저항’ (resistance) 등이 높은 유사도를 나타냈다. ‘움직임’ (motion)의 경우 ‘관절가동범위’ (ROM), ‘움직임’ (movement), ‘돌림’ (rotation), ‘벌림’ (abduction) 등과 같이 어깨 관절의 움직임과 관련된 용어들이 높은 유사도를 나타냈다. ‘상지’ (upper)는 ‘말단’ (extremity), ‘수족’ (limb), ‘운동’ (motor) 등이 높은 유사도를 나타냈다.

4. t-SNE

Fig. 3은 워드-임베딩 방식으로 300차원 공간으로 임베딩 시킨 단어들을 t-SNE 방식을 통해 2차원 공간에 나타낸 그래프이다. ‘어깨’ (shoulder), ‘재활’ (rehabilitation) 및 전체 단어 중 가장 높은 빈도수를 보인 10개의 키워드들의 위치와 거리가 가까운 단어들을 표기하였다. Table 3은 이들 단어들과 거리가 가까운 10개의 단어들의 목록과 거리를 제시하였다. 이를 통해, 어깨 재활 관련 연구에서 이들 키워드 간의 연구적 연결성을 시각적으로 살펴볼 수 있다.

어깨 재활 관련 연구들의 키워드들은 크게 다섯개의

그룹으로 구분되는 것으로 나타났다. 각각의 그룹을 A그룹, B그룹, C그룹, D그룹, E그룹으로 구분하였다. A그룹은 ‘어깨’ (shoulder)를 중심으로 이와 가장 근접한, ‘팔꿈치’ (elbow), ‘팔’ (arm), ‘상지’ (upper), ‘손’ (wrist) 그리고 ‘움직임’ (motion) 등의 단어들로 구분되었고, B그룹은 ‘재활’ (rehabilitation)을 중심으로 ‘프로그램’ (program), ‘치료’ (treatment), ‘관리’ (management), ‘돌봄’ (care), ‘물리치료’ (physiotherapy) 등의 단어들이 가까운 거리에 위치하였다. C그룹은 ‘점수’ (score)를 중심으로 ‘결과’ (outcome), ‘척도’ (scale), ‘장애’ (disability), ‘시각통증척도’ (VAS), ‘미국 어깨 및 팔꿈치 외과 의사 점수’ (ASES) 등의 단어들이 가까운 거리에 위치하였으며, D그룹은 ‘환자’ (patient)를 중심으로 ‘대상자’ (subject), ‘케이스’ (case)와 같은 단어들이 가까운 거리에 위치하고 있었다. E그룹은 ‘손상’ (injury)을 중심으로 ‘외상’ (trauma), ‘운동선수’ (athlete), ‘스포츠’ (sport), ‘병변’ (lesion), ‘과사용’ (overuse) 등의 단어들이 가까운 거리를 가지는 것으로 구분되었다.

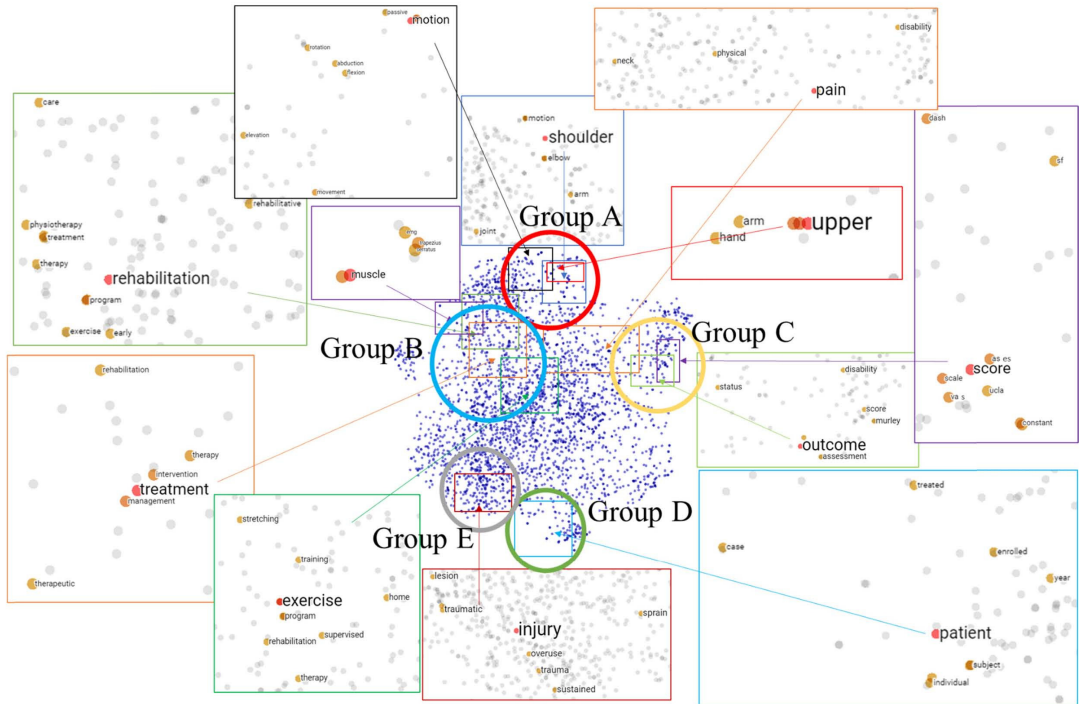


Fig. 3. t-SNE 2D visualization of 12 keywords.

Table 3. Distance of 12 Keywords in t-SNE

Rank	shoulder	rehabilitation	patient	Pain	treatment	muscle	score	outcome	injury	exercise	motion	upper
1	elbow	program	subject	symptom	management	trapezius	ases	measure	trauma	strengthening	rom	extremity
2	.636	.521	.631	.560	.485	.537	.492	.635	.608	.501	.489	.435
3	pain	therapy	case	disability	intervention	emg	scale	functional	athlete	stretching	range	limb
4	.643	.593	.636	.610	.516	.594	.514	.687	.646	.597	.538	.462
5	glenohumeral	management	treated	complaint	therapy	activation	dash	score	sport	program	movement	muscle
6	.670	.643	.639	.638	.602	.601	.535	.714	.691	.609	.574	.660
7	joint	care	month	shoulder	conservative	deltoid	vas	status	lesion	resistance	passive	motor
8	.687	.648	.677	.643	.633	.607	.545	.730	.704	.613	.633	.664
9	arm	treatment	individual	intensity	rehabilitation	infrapinatus	constant	disability	overuse	training	rotation	lower
10	.695	.649	.678	.661	.649	.609	.572	.735	.712	.621	.652	.681
11	also	rehabilitative	year	problem	operative	serratus	murley	improvement	dislocation	programme	abduction	arm
12	.697	.671	.680	.712	.692	.637	.590	.742	.719	.637	.683	.712
13	motion	physiotherapy	people	chronic	treat	musculature	sf	follow	incidence	supervised	elevation	hand
14	.715	.678	.683	.717	.695	.642	.603	.742	.727	.696	.703	.726
15	wrist	early	participant	painful	nonoperative	upper	respectively	primary	player	therapy	shoulder	trunk
16	.723	.688	.690	.729	.695	.660	.612	.744	.736	.701	.715	.735
17	range	protocol	treatment	headache	patient	electromyographic	ucla	reported	competition	activity	flexion	ue
18	.731	.691	.695	.743	.695	.678	.622	.747	.740	.711	.718	.766
19	extremity	programme	enrolled	life	treated	contraction	index	clinical	disorder	rehabilitation	mobility	impairment
20	.742	.699	.702	.745	.696	.685	.634	.752	.747	.711	.734	.767

IV. 고 찰

본 연구는 텍스트 마이닝 기법을 활용하여 1971년부터 2021년 3월까지 ‘어깨’와 ‘재활’을 키워드를 포함하고 있는 해외 연구 10,121편의 초록을 바탕으로 연구 동향을 분석하였다. 어깨 재활 관련 연구들의 연구 동향과 특징은 다음과 같이 볼 수 있을 것이다.

첫째, 어깨 재활 관련 연구들은 지난 50년 동안 굉장히 많은 증가를 나타냈다. 출판된 어깨 재활 관련 연구들을 10년의 단위로 나누어 살펴보면 매 10년마다 2배를 넘는 수치로 연구 수가 증가하는 것으로 나타났다. 특히 지난 10년동안 매해 발표되는 어깨 재활 관련 연구의 수도 지속적으로 증가한 것으로 나타났다.

둘째, 빈도수를 바탕으로 수집한 키워드들을 통해 어깨 재활 관련 연구들은 ‘환자’ (patient), ‘통증’ (pain), ‘치료’ (treatment)를 중심으로 수행되어 왔다는 것을 알 수 있다. 이러한 연구들은 ‘움직임’ (range, motion), ‘관절’ (joint), ‘근육’ (muscle) 등에 대한 증재들을 바탕으로 이루어졌을 것으로 분석해 볼 수 있다. 연도별로 키워드를 살펴보았을 때 시간이 흐름에 따라 새로운 키워드가 등장하기도 하거나 키워드들의 빈도수가 변하기도 하였다. 1981년도부터는 ‘활동’ (activity)이라는 단어를 통해 근전도(muscle activity)와 관련된 연구들과 어깨 병변을 가진 환자들의 활동(activity)에 관한 분석 연구들이 본격적으로 수행되었을 것으로 분석해볼 수 있다. 이에 대한 예로, 주파수 스펙트럼 분석 연구들과 같이 근전도 장비를 이용한 근피로도 연구들이 이 시기에 활발하게 진행되기도 하였다[27-30]. 2001년부터는 ‘점수’ (score)라는 단어를 통하여 어깨 재활에 관한 평가와 연구들이 점수화를 이용한 정량적 평가 방법들로 많이 수행되었음을 확인해 볼 수 있다. 실제로, 1987년 Constant scale이 어깨 관절 평가에 가장 보편적으로 사용되기 시작한 이래로 미국 어깨관절 외과 학회에서는 1993년 ASES 표준양식을 개발하였고, 이후 미국 골관절 수술 협회에서는 1999년 DASH 매뉴얼을 개발하여 출간하는 등 어깨 평가 도구들에 관한 연구들이 본격적으로 수행되었다[31-33]. 이와 함께, ‘운동’ (exercise)이라는 단어가 나타나기 시작하였고 점차적으로 그 빈도

수가 증가하기 시작하였다. 이와 비슷한 시기에 세계 보건 기구(World Health Organization; WHO)는 심혈관 질환, 암, 당뇨, 기타호흡기질환이 전 인구 사망률의 60%를 차지하고 이에 대한 위험인자는 운동 부족이라고 경고할 정도로 건강한 삶을 증진 시키기 위하여 운동의 중요성을 강조하였다[34]. 이를 통해, 어깨 재활 영역에서도 ‘운동’과 관련된 주제들에 대한 관심이 증가했을 것이고 ‘운동’이라는 단어가 연구에서 더 많이 사용되었을 것이다. 2011년도부터 시작된 연구에서는 ‘결과’ (outcome)라는 단어의 빈도가 증가하였다. 이는 이전에 나타났던 ‘점수’ (score) 단어와 함께 정량적 평가들을 이용한 연구들이 더욱 많이 증가했다는 사실을 유추해 볼 수 있다.

셋째, 워드-임베딩 분석을 통하여 ‘어깨’ (shoulder), ‘재활’ (rehabilitation), 그리고 상위 키워드 10개 단어들과 높은 유사도를 가지고 있는 단어들을 분석하였다. 그리고 이러한 단어들을 t-SNE 방법을 이용하여 어깨 재활 관련 연구들의 키워드 및 유사도가 높은 단어들을 시각화 하였으며 이를 바탕으로 5개의 그룹별로 구분하여 그 특징을 분석하였다. 각 그룹은 중심 키워드를 바탕으로 구성이 되었는데 그 중심 키워드는 각각 ‘어깨’ (shoulder), ‘재활’ (rehabilitation), ‘점수’ (score), ‘환자’ (patient), ‘손상’ (injury)이었다. A그룹에서 ‘어깨’ 키워드와 근접한, ‘팔꿈치’ (elbow), ‘팔’ (arm), ‘상지’ (upper), ‘손’ (wrist) 등의 단어들을 통해 어깨가 상지의 주변 부위와 연관성이 높게 연구되고 있는 것으로 나타났다. 또한 ‘움직임’ (motion) 단어를 통하여 ‘어깨’와 ‘움직임’이 높은 연관성으로 연구되는 온 것으로 나타났다. B그룹은 ‘재활’ (rehabilitation) 키워드를 중심으로 ‘프로그램’ (program), ‘치료’ (treatment), ‘관리’ (management), ‘돌봄’ (care) 단어들이 밀접한 연관성을 가지는 것으로 나타났다, 이와 더불어 ‘물리치료’ (physiotherapy), ‘운동’ (exercise), ‘근력강화’ (strengthening) 등의 단어들을 통해 ‘치료’, ‘증재’에 대한 방법으로 ‘물리치료’나 ‘운동’ 및 ‘근력강화’와 같은 방법들이 사용되었을 것으로 분석해 볼 수 있다. C그룹은 ‘점수’ (score) 키워드를 중심으로 ‘결과’ (outcome), ‘척도’ (scale), ‘장애’ (disability) 등의 단어들과 밀접한 연관성을 가지며, 이러한 단어들에

대한 요소로 적용될 수 있는 ‘시각통증척도’ (VAS), ‘미국 어깨 및 팔꿈치 외과 의사 점수’ (ASES)와 같은 평가 방법들이 높은 연관성을 나타냈다. D그룹은 ‘환자’ (patient)를 중심으로 ‘대상자’ (subject), ‘케이스’ (case)와 같은 연구 대상자들의 특성과 관련된 단어들이 높은 연관성을 가지는 것으로 나타났다. 그리고, E그룹은 ‘손상’ (injury)을 중심으로 ‘외상’ (trauma), ‘병변’ (lesion), ‘과사용’ (overuse) 등 단어들이 높은 연관성을 가지는 것으로 나타났다. 이와 함께, ‘운동선수’ (athlete), ‘스포츠’ (sport)와 같은 단어들을 통해 해당 단어들이 의학 분야뿐만 아니라 스포츠와 같은 다양한 분야와 폭넓게 연구가 수행되었다는 사실을 분석해볼 수 있다.

이 연구의 한계점으로는 다음과 같다. 첫째, 이 연구에서는 검색어를 ‘어깨’ (shoulder)와 ‘재활’ (rehabilitation)을 포함하는 논문을 분석하였기에 재활 관련 연구를 대표한다고 볼 수 없다. 둘째, 의학 서지 데이터베이스인 MEDLINE 자료들만 사용하였기에 이에 포함되지 않은 다른 데이터베이스의 연구들은 제외되었다. 셋째, 이 연구에서는 1971년부터 2021년 3월까지의 논문 초록 데이터를 사용하여 주요 단어들의 관계성을 모형화 하였으나 시간의 흐름에 따른 주요 단어들의 관계성이나 트렌드에 대한 분석은 부족하다.

V. 결론

본 연구는 텍스트 마이닝 기법을 활용하여 ‘어깨’ (shoulder), ‘재활’ (rehabilitation) 관련 해외 학술지 논문들의 초록을 구성하는 주요 단어들과 이들의 관계성을 모델링한 첫 연구이다. 키워드 분석을 통해 ‘어깨’, ‘재활’ 관련 연구들은 ‘환자’ (patient), ‘통증’ (pain), ‘치료’ (treatment) 단어들을 가장 많이 사용하였다는 사실을 알 수 있었으며, 텍스트 마이닝을 이용한 빅데이터 분석방법을 통해 키워드들과 관련성이 높은 단어들에 대한 파악이 가능하였다. 또한 5가지의 그룹화를 통하여 ‘어깨’, ‘재활’ 관련 연구들의 동향을 파악할 수 있었으며 이는 앞으로 수행될 어깨 재활 연구들에 대한 주제적 다양성을 높이는데 도움을 줄 수 있을 것이다.

References

- [1] Luime JJ, Koes BW, Hendriksen IJM, et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol*. 2004;33(2):73-81.
- [2] Oh TY, Ha JY, Lee EJ, et al. The effects of badminton on the shoulder and its correlation with the shoulder pain and disability index. *J Korean Phys Ther*. 2013;25(1):29-35.
- [3] Triffitt PD. The relationship between motion of the shoulder and the stated ability to perform activities of daily living. *J Bone Jt Surg*. 1998;80(1):41-6.
- [4] Kim J-H, Kwon O-Y, Hwang U-J, et al. Comparison of the Shoulder External Rotator Strength and Asymmetry Ratio Between Workers With and Without Shoulder Impingement Syndrome. *J Strength Cond Res*. 2019;
- [5] Møller SP, Brauer C, Mikkelsen S, et al. Risk of subacromial shoulder disorder in airport baggage handlers: combining duration and intensity of musculoskeletal shoulder loads. *Ergonomics*. 2018;61(4):576-87.
- [6] Curl LA, Warren RF. Glenohumeral Joint Stability: Selective Cutting Studies on the Static Capsular Restraints. *Clin Orthop Relat Res*. 1996;330:54-65.
- [7] Van der Windt DA, Koes BW, de Jong BA, et al. Shoulder disorders in general practice: incidence, patient characteristics, and management. *Ann Rheum Dis*. 1995;54(12):959-64.
- [8] Mitchell C, Adebajo A, Hay E, et al. Shoulder pain: diagnosis and management in primary care. *Bmj*. 2005;331(7525):1124-8.
- [9] Yoshimizu M, Teo AR, Ando M, et al. Relief of chronic shoulder and neck pain by electro-acupuncture and transcutaneous electrical nervous stimulation: A randomized crossover trial. *Med Acupunct*. 2012;24(2): 97-103.
- [10] Ide J, Maeda S, Yamaga M, et al. Shoulder-strengthening exercise with an orthosis for multidirectional shoulder instability: quantitative evaluation of rotational shoulder strength before and after the exercise program. *J Shoulder Elb Surg*. 2003;12(4):342-5.

- [11] Nadler M, Pauls MMH. Shoulder orthoses for the prevention and reduction of hemiplegic shoulder pain and subluxation: systematic review. *Clin Rehabil.* 2017;31(4):444-53.
- [12] Andersen JH, Kaergaard A, Frost P, et al. Physical, psychosocial, and individual risk factors for neck/shoulder pain with pressure tenderness in the muscles among workers performing monotonous, repetitive work. *Spine.* 2002;27(6):660-7.
- [13] Roquelaure Y, Mariel J, Fanello S, et al. Active epidemiological surveillance of musculoskeletal disorders in a shoe factory. *Occup Environ Med.* 2002;
- [14] Camargo PR, Haik MN, Mattiello-Rosa SM, et al. Pain in workers with shoulder impingement syndrome: an assessment using the DASH and McGill pain questionnaires. *Braz J Phys Ther.* 2007;11(2):161-7.
- [15] Hill L, Collins M, Posthumus M. Risk factors for shoulder pain and injury in swimmers: A critical systematic review. *Phys Sportsmed.* 2015;43(4):412-20.
- [16] Andersson SH, Bahr R, Clarsen B, et al. Preventing overuse shoulder injuries among throwing athletes: a cluster-randomised controlled trial in 660 elite handball players. *Br J Sports Med.* 2017;51(14):1073-80.
- [17] Reeser JC, Joy EA, Porucznik CA, et al. Risk factors for volleyball-related shoulder pain and dysfunction. *Pm&r.* 2010;2(1):27-36.
- [18] Jun S. A Big Data Preprocessing using Statistical Text Mining. *J Korean Inst Intell Syst.* 2015;25(5):470-6.
- [19] Kim S, Lee Y, Shin J, et al. Text Mining for Economic Analysis. *J Korean Econ Anal.* 2020;26(1):1-85.
- [20] Cho S, Kim S. Finding Meaningful Pattern of Key Words in IIE Transactions Using Text Mining. *J Korean Inst Ind Eng.* 2012;38(1):67-73.
- [21] Jo E-J, Kim K-Y. Analysis of Occupational Therapy-Related Keywords Using Text Mining. *J Korean Soc Occup Ther.* 2020;28(3):39-51.
- [22] Shin M-S, Cho K-W. Analysis on Topic Modeling and Trend of Journal of Speech-Language & Hearing Disorders using Text Mining: (2002~2018). *J Speech-Language Hear Disord.* 2019;28(3):81-91.
- [23] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality. *Neural Inf Process Syst.* 2013;3111-9.
- [24] Heu J. Korean Language Clustering using Word2Vec. *J Inst Internet, Broadcast Commun.* 2018;18(5):25-30.
- [25] Zhu Y, Yan E, Wang F. Semantic relatedness and similarity of biomedical terms: examining the effects of recency, size, and section of biomedical publications on the performance of word2vec. *BMC Med Inform Decis Mak.* 2017;17(1):1-8.
- [26] Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res.* 2008;9(11):2579-605.
- [27] Petrofsky JS, Lind AR. The influence of temperature on the amplitude and frequency components of the EMG during brief and sustained isometric contractions. *Eur J Appl Physiol Occup Physiol.* 1980;44(2):189-200.
- [28] Basmajian J V. EMG signal amplitude and force. *Muscles Alive.* 1985.
- [29] De Luca CJ. Myoelectrical manifestations of localized muscular fatigue in humans. *Crit Rev Biomed Eng.* 1984;11(4):251-79.
- [30] Fugl-Meyer AR, Gerdle B, Eriksson BE, et al. Isokinetic plantar flexion endurance. Reliability and validity of output/excitation measurements. *Scand J Rehabil Med.* 1985;17(2):47-52.
- [31] Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. *Clin Orthop Relat Res.* 1987;(214):160-4.
- [32] Richards RR, An K-N, Bigliani LU, et al. A standardized method for the assessment of shoulder function. *J Shoulder Elb Surg.* 1994;3(6):347-52.
- [33] Germann G, Wind G, Harth A. The DASH (Disability of Arm-Shoulder-Hand) Questionnaire—a new instrument for evaluating upper extremity treatment outcome. 1999;31(3):149-52.
- [34] Organization WH. The world health report 2002: reducing risks, promoting healthy life. World Health Organization; 2002.