

# Namu Park

1212 Harrison Street, Seattle, WA, 98109

Cell: (206) 992-9325

Email: namupark@yonsei.ac.kr

## RESEARCH INTERESTS

- **Data Science:** Medical Data Science, Information Retrieval, Social Text Mining
- **Natural Language Processing:** NLP for Clinical Informatics (Narrative Radiology Reports, Electronic Health Records), Distributed Representation, Machine Translation
- **Machine Learning:** Artificial Intelligence, Machine Learning for Electronic Health Records, Representation Learning (Triplet Network), Unsupervised Learning

## EDUCATION

**University of Washington**, Seattle, Washington

Advisor: Professor Meliha Yetisgen

- Ph.D. in Biomedical & Health Informatics (2021 - 2026)

**Yonsei University**, Seoul, South Korea (CGPA: 3.98/4.00)

Advisor: Professor Min Song

- M.S. in Digital Analytics (2020)
- Master's thesis: **Information Extraction from Unstructured Medical Text using Pseudo-label-based Semi-supervised Learning**

**Sogang University**, Seoul, South Korea (CGPA: 3.59/4.00, Triple major, *Magna Cum Laude*)

- B.S. in Convergence Software (2019)
- B.E. in Economics (2019)
- B.A. in French Culture (2019)

### Courses related to Research Interests:

Computer Science	Data Science	Math / Statistics / Analytics
- Data Structures <sup>†</sup>	- Machine Learning <sup>††</sup>	- Linear Algebra <sup>†</sup>
- Python <sup>†</sup>	- Database Management <sup>††</sup>	- College Mathematics <sup>†</sup>
- Java Programming <sup>†</sup>	- Text Mining <sup>††</sup>	- Mathematical Economics <sup>†</sup>
- C Programming <sup>†</sup>	- Data Mining <sup>††</sup>	- Econometrics <sup>†</sup>
- Operating Systems <sup>†</sup>	- Big Data Parallel Processing <sup>††</sup>	- Biostatistics <sup>††</sup>
- Algorithms <sup>†</sup>	- Artificial Intelligence and Deep Learning <sup>††</sup>	- Basics of Big Data Analytics <sup>††</sup>
- Database Systems <sup>†</sup>	- Advanced Machine Learning <sup>††</sup>	- Big Data Statistical Analytics <sup>††</sup>
- Capstone Design <sup>†</sup>	- Natural Language Processing and Deep Learning <sup>††</sup>	- Practical Big Data Analytics <sup>††</sup>

† : Undergraduate, †† : Graduate

## PUBLICATIONS

- **Prediction of Lung Cancer TNM Staging in PET-CT Clinical Notes**  
Namu Park, Hyung-Jun Park, Min Song, Chang-Min Choi (in preparation)
- **Are we there yet? Analyzing scientific research related to COVID-19 drug repurposing**  
Namu Park, Hyeyoung Ryu, Ying Ding, Qi Yu, Yi Bu, Qi Wang, Jeremy J. Yang, Min Song. *18th International Conference on Scientometrics and Informetrics (ISSI)*, 2021.
- **Analyzing knowledge entities about COVID-19 using entitymetrics**  
Qi Yu, Qi Wang, Yafei Zhang, Chongyan Chen, Hyeyoung Ryu, Namu Park, ... , Yi Bu. *Scientometrics*, 2021.
- **Information Extraction from Unstructured Medical Text using Pseudo-label-based Semi-supervised Learning**

Namu Park, Min Song

*Yonsei University Graduate School Dissertations, 2020.*

- **A Monte Carlo Search-based Triplet Sampling Method for Learning Disentangled Representation of Impulsive Noise on Steering Gear**  
Seok-Jun Bu, Namu Park, Gue-Hwan Nam, Jae-Yong Seo, Sung-Bae Cho.  
*IEEE, International Conference on Acoustics, Speech and Signal Processing, 2020.*  
(Virtual Presentation Speaker)
- **Data Augmentation using Empirical Mode Decomposition on Neural Networks to Classify Impact Noise in Vehicle**  
Gue-Hwan Nam, Seok-Jun Bu, Namu Park, Jae-Yong Seo, Hyeon-Cheol Jo, Won-Tae Jeong.  
*IEEE, International Conference on Acoustics, Speech and Signal Processing, 2020.*
- **Classifying Impact Noise in Car Steering Gear using Mel-spectrogram-based Convolutional-Recurrent Neural Network**  
Namu Park, Seok-Jun Bu, Sung-Bae Cho.  
*Korea Software Congress, 2019.*

## RESEARCH EXPERIENCE

---

### Research Assistant

*BioNLP Lab. at University of Washington*

**Sept. 2021 – Present**

Advisor: Professor Meliha Yetisgen

- **Large Scale Clinical and Economic Impact Analysis of Potentially Malignant Incidental Findings in Radiology Reports**
  - Joint research with Professor Martin Gunn, *Department of Radiology, University of Washington*
  - **Research goal: Extracting meaningful information in radiology reports to find cancer-related “incidentalomas”, which indicate incidental findings that are potentially malignant**
  - Leading the annotation phase while co-working with 4 medical students. Labels such as “Lesion”, “Indication”, “Medical Problem” are used.

### Researcher

*Asan Institute for Life Sciences, Asan Medical Center*

**Sept. 2020 – Aug. 2021**

Advisor: Professor Chang-Min Choi

- **Development of the AI-based Cardiopulmonary Disease Prediction System using Biological Ring Sensors**
  - Government-funded research supported by the Cross-Ministry Korea Medical Device Development Fund
  - **Research goal: Providing emergency warnings and inducing visits to medical institutions through machine learning-based analysis in case of frequent unusual symptoms or high probability of cardiopulmonary disease**
  - Implemented the neural network model that can predict pneumonia through the analysis of cardiopulmonary features (e.g. SpO2 and pulse rate)

### Research Assistant

*Deep Text Lab. at Yonsei University*

**Mar. 2019 – Aug. 2020**

Advisor: Professor Min Song

- **Information Extraction in PET-CT Clinical Notes related to Lung Cancer**
  - Joint research with Professor Chang-Min Choi, *Department of Pulmonology and Critical Care Medicine, Asan Medical Center, College of Medicine, University of Ulsan*
  - **Research goal: Extracting accurate information in PET-CT radiology reports such as primary lung cancer area and metastasis area, which are critical in determining the TNM staging of lung cancer**

- Collected 1,242 papers in PubMed and 2,318 papers in PubMed Central, which were all related to lung cancer and developed Lung Cancer Spell Checker (LCSC), a typo corrector and abbreviation standardization tool used for preprocessing
- Applied *FastText*, a character-level embedding method known to be robust in typos, for efficient representation of each word into a 100-dimensional vector
- Solved the lack of labeled data problem by implementing pseudo-label-based semi-supervised learning, which yields the most probable label for each data
- Wrote an instruction manual to facilitate the use of Python-based information extraction model for medical experts without any programming knowledge
- **Consensus Analysis of Drug Repurposing Literatures for COVID-19**
  - Joint research with Professor Ying Ding, *School of Information, University of Texas, Austin*
  - **Research goal: Observation of recent drug repurposing literature related to COVID-19 to check whether a potential consensus regarding proposed drugs exists or not**
  - Transformed drug repurposing papers, which were collected via PubMed, into a 100-dimensional vector in the *BioCOVID-BERT* embedding space
  - Plotted each document vector in 2-dimensional space using Principal Component Analysis to find the most similar documents in terms of embedded vectors
  - Examined the conspicuous consensus in COVID-19 drug repurposing papers with the help of document clustering using Term Frequency Inverse Document Frequency (TF-IDF)
  - Compared with indistinct consensus, which was based on degree centrality, betweenness centrality, and closeness centrality of drug entities
- **Automatic Translation of Affiliations and Author Names in Research Papers using Attention and Seq2Seq**
  - Joint research with *College of Medicine, Yonsei University*
  - **Research goal: English-to-Korean translation of affiliation and author name to facilitate document management, especially for scholarly manuscripts in the medical domain**
  - Collected 7,100 first names and 113 last names and used their combinations as train data for character-level Long Short-Term Memory Seq2Seq model
  - Collected 200 department names within the College of Medicine from websites of 18 universities and included 51,644 medical terms to generate robust train data for Attention-based neural machine translation model
  - Combined JAVA-based information extraction model and Python-based translation model using JPype
  - Developed a Graphical User Interface (GUI) using PYQT5 and wrote an instruction manual for the GUI, targeting users who are not familiar with computer programming

**Research Assistant**

*Soft Computing Lab. at Yonsei University*

**Aug. 2019 – Mar. 2020**

Advisor: Professor Sung-Bae Cho

- **Detection of Potential Toxic Clause**
  - Joint research with *Samsung Engineering IT innovation team (Samsung)*
  - **Research goal: Classification of toxic clauses to improve risk detection of potential bad overseas contracts**
  - Developed a risk detection model using neural network architecture and enabled automatic update every time new data comes in
  - Developed the Python code to facilitate the use of GPUs within the company and gathered data from both the local database and DB server (the connection between Python code and database by implementing SQL queries in Python)
  - Visited *Samsung Engineering* office twice to give a demonstration of the toxic clause detection module

- **Deep Learning-based Gear Noise Classification**
  - Joint research with *Hyundai Mobis NVH team (Hyundai Motor Group)*
  - **Research goal: Classification of 4 similar types of steering gear noise to indicate possible defects in different parts of the vehicle**
  - Proposed a system for inspection of impact noise on steering gear based on neural network architecture and log Mel-Spectrogram
  - Extracted spatio-temporal features with Convolutional-Recurrent Neural Network (1D-Temporal CNN-LSTM)
  - Compared classification results with different feature extraction methods (Raw waveform, Short-Term Fourier Transform, Mel-Frequency Cepstral Coefficient)
  - Discovered that the MFCC-based model had the best classification accuracy due to efficient low-frequency sound extraction
  - Showed high performance when practically applied to the real vehicle

#### Research Trainee

Jun. 2018 – Sept. 2018

*Big Data X Campus, Government of the Republic of Korea*

- Research-oriented data science summer school for undergraduate students (participation funded by the *Government of the Republic of Korea*)
- Learned basic machine learning and deep learning theories
- Practical programming exercise using Python, Tensorflow, Apache Spark, Hadoop framework
- **Deep Learning-based Disease Prediction without Blood-gathering**
  - Final group project for the training program
  - **Research goal: Reducing the need for blood-gathering and allowing the public to receive healthcare in an easier and convenient way**
  - Gathered data on Korean body measurements and medical check-up reports
  - Used weight, height, waist measurement, age, smoking as inputs and blood sugar level, blood pressure, total cholesterol, HDL cholesterol, LDL cholesterol, hemoglobin as outputs
  - Based on mentioned outputs, showed the probability of related diseases such as high blood pressure and diabetes

## TEACHING EXPERIENCE

---

#### Lecturer

Oct. 2019

*Korea Industrial Technology Association*

- Lecture on Recurrent Neural Network and Long Short-Term Memory (targeting researchers not familiar with machine learning)
- Python tutorial on deep learning using Tensorflow, Keras
- Instructed hands-on project on text generation using Wikipedia dataset

#### Teaching Assistant

Spring 2019

*Department of Digital Analytics, Yonsei University*

- Course: Database Management
- Helped students having difficulties in database theories
- Prepared tutorials to instruct the application of various SQL queries in MySQL
- Special lecture on basic and advanced SQL queries
- Special lecture on basic database management theory focused on Relational Database (RDB)

#### Teaching Assistant

Spring 2019

*Department of Digital Analytics, Yonsei University*

- Course: Big Data and Knowledge Discovery
- Provided supplementary information on big data theory (necessity of parallel computing)
- Proctored 5 quizzes on data warehouse and big data application

## **SCHOLARSHIPS / FELLOWSHIPS**

---

- UW Graduate School Top Scholar Top Off Fellowship – University of Washington (2020)
- UW Provost Top Scholar Top Off Fellowship – University of Washington (2020)
- UW Research Assistant Scholarship (2020)
- Yonsei Digital Analytics Teacher Assistant Scholarship (2019, 2020)
- Higher Education Innovation Team Social Innovation Activity Scholarship (2019)
- Samsung Convergence Software Course Scholarship - **Academic Excellence** (2018)
- Sogang Honors Scholarship - **Academic Excellence** (2017)
- Government of the Republic of Korea Funding Scholarship (2016)
- Sogang SALANG Scholarship (2013, 2016, 2017, 2018)

## **TECHNICAL SKILLS AND CERTIFICATES**

---

### **Technical Skills**

- Programming Language - Python, C, C++, Java, R
- Machine Learning - Tensorflow, Keras, PyTorch
- SQL (Structured Query Language) - MariaDB, Oracle SQL developer
- Big Data Analytics - Hadoop, Spark
- Others - Google Firebase, Android Studio, Django Web, HTML, LaTeX

### **Certificates**

- Excellence Award by commissioner of the *Seoul Metropolitan Police Agency* for performing CPR and saving life of a pedestrian
- Biology Meets Programming (Coursera certificate no. RQE7WKMXMSEB)
- SCSC (Samsung Convergence Software Course) certificate
- Big Data X Campus certificate

## **OTHER INFORMATION**

---

- Honorably discharged as Sergeant **Feb. 2014 – Nov. 2015**
  - Military duty as auxiliary police (interpretation of English/French/Korean, maintenance of public security)
  - Squad leader of Tourist Police Auxiliary Police
  - Participated in 2015 Gwangju Summer Universiade as interpreter and security guard
- Member of Sogang University Basketball Team **Mar. 2013 – Present**
  - Varsity basketball player of Sogang University at 2017 Sogang-Sophia Festival of Exchange
  - 2<sup>nd</sup> place in Kyonggi University Basketball Tournament
- Lived 2 years in Montreal, Canada **Apr. 2004 – Mar. 2006**
- Lived 2 years in Paris, France **Mar. 1997 – Jun. 1999**