

자동차 주행 빅데이터를 이용한 신뢰성 분석 인공지능 플랫폼의 개발

김 중 호¹ · 박 화 규^{2*}¹경성대학교 경영학과 부교수 ^{2*}순천향대학교 보건행정경영학과 교수 (교신저자)

Development of Reliability Analysis AI Platform based on Vehicle Driving Big Data

Jong-Ho Kim¹ · Hwa-Kyu Park^{2*}¹Associate Professor, Department of Business Administration, Kyungsung University, Busan 48434, Korea^{2*}Professor, Department of Health Administration, Soon Chun Hyang University, Choongnam 31538, Korea

[요 약]

완성차와 부품업체는 제품 출시 전후 주행시험을 상시적으로 진행하고 있다. 이 과정에서 다양한 센서를 통해 수집한 빅데이터를 보유하고 있지만 좀처럼 제품의 개선과 혁신으로 연결하지 못하고 있다. 자동차 주행 성능 테스트가 워낙 다양한 모드와 환경에서 수행되고 있고 많은 변수와 데이터셋이 생성되기 때문에 엔지니어들은 방대한 주행 빅데이터를 분석해서 최선의 결론을 찾는데 큰 어려움을 겪고 있다. 본 연구의 목표는 엔지니어들이 이러한 어려움을 극복하고 방대한 주행 빅데이터를 분석하여 성능평가, 고장진단, 신뢰성 분석을 효과적으로 수행할 수 있도록 지원하는 신뢰성 분석 시스템을 개발하는 것이다. 이를 위해 자동차 신뢰성 평가 프로세스를 정립하고 시스템의 기능을 도출한 후 타겟 아키텍처를 설계하였다. 개발과정에서 프론트엔드의 요구사항과 백엔드 데이터베이스 설계 요구사항을 함께 수용하여 웹 프로토콜을 이용하여 실행되는 함수를 애플리케이션 레이어 내에서 디자인하고 API로 구현하는 API First 기법을 적용하였다. 개발된 플랫폼은 내연기관, 하이브리드, 전기자동차 등 다양한 차종을 대상으로 실시되는 엔진대상, 스피드웨이 주행시험, 실주행 시험 과정에서 발생하는 빅데이터의 수집, 전처리, 분석, 시각화 전 주기를 지원한다.

[Abstract]

Automobiles and parts makers are constantly conducting driving tests before and after product launches. They have big data collected through various sensors in this process, but it is rarely connected to product improvement and innovation. Because automobile driving performance tests are conducted in so many different modes and environments, and many variables and datasets are generated, engineers face great difficulties finding conclusions by analyzing huge driving big data. The goal of this study is to develop reliability analysis system that supports engineers to overcome these difficulties and effectively perform performance evaluation, fault diagnosis, and reliability analysis by analyzing vast driving big data. To this end, the vehicle reliability evaluation process was established, the system function was derived, and the target architecture was designed. Adopting API first approach we designed APIs which take the requirements of the front-end and the back-end layers. Next, we developed the functions using the web protocol within the application layer based on the design artifacts. The system supports the entire cycle of data collection, pre-processing, analysis, and visualization in the engine-only test, speedway driving test, and actual road test for various vehicle types such as internal combustion engines, hybrids, and electric vehicles.

색인어 : 빅데이터, 차량주행, 신뢰성, 플랫폼, 인공지능**Keyword** : Big Data, Vehicle Driving, Reliability, Platform, Artificial Intelligence<http://dx.doi.org/10.9728/dcs.2022.23.12.2425>

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.

Received 20 September 2022; Revised 18 October 2022

Accepted 28 November 2022

***Corresponding Author; Hwa-Kyu Park**

Tel: +82-41-530-1740

E-mail: hkpark1@sch.ac.kr

1. 서론

완성차와 부품업체는 제품 출시 전후 주행시험을 상시적으로 진행하고 있으며 이 과정에서 다양한 센서를 통해 수집한 빅데이터를 보유하고 있다. 자동차 제조기업들은 연비, 배기가스 규제를 만족시키고 소비자의 기대 욕구에 부응하기 위하여 다양한 환경에서 자동차 주행성능시험을 진행하고 있으며 부품업체 역시 실제와 동일한 조건으로 부품의 성능 및 내구 시험을 진행하면서 많은 실험 데이터를 보유하고 있다. 그러나 자동차 기업들은 보유하고 있는 빅데이터의 분석을 통해 제품의 개선과 혁신으로 연결하는 데 많은 어려움을 겪고 있다. 자동차 주행시험이 워낙 다양한 모드와 환경에서 수행되고 있고 많은 변수와 데이터셋이 생성되기 때문에 엔지니어들이 방대한 주행 빅데이터를 분석해서 최선의 결론을 찾는 데 많은 한계가 있기 때문이다.

주행시험 빅데이터 분석 과정에서 연고자 하는 핵심은 제품의 신뢰성이다. ‘신뢰성’이 자동차 제조기업의 시장경쟁력 및 경영성과와 직결됨에 따라 자동차 주행성능 분석에서도 전통적인 품질관리에서 신뢰성 분석으로 빠르게 전환되고 있다. 일반적으로 신뢰성이란 어떤 장비, 부품, 소재가 주어진 시간 동안 특정 사용조건에서 요구 기능을 만족스럽게 유지시킬 수 있는 가능성(확률)을 의미하고 신뢰성 분석은 고장의 예측, 예방, 관리를 다루는 분야이다[1]. 주어진 시간은 기능 수행을 위해 설정된 목표 시간이며 경쟁제품과의 비교, 교체 정책 등을 종합적으로 고려하여 설정된다. 사용조건으로서 자동차의 경우 외기 온도, 습도, 에어컨 작동 여부, 주행시간, 주행장소, 탑승인원 (체중합계) 등이 있다.

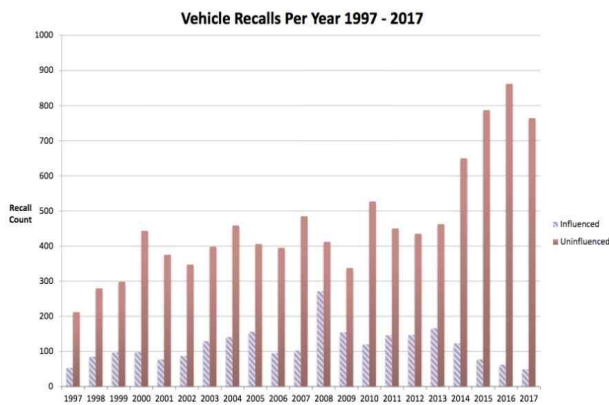


그림 1. 자동차 리콜 추이 (출처 : Zero Defect Manufacturing)
 Fig. 1. The trend of Car Recall (Source : Zero Defect Manufacturing)

소비자의 불신, 리콜 건수의 증가, 천문학적 리콜비용 등으로 신뢰성은 자동차 기업의 시장경쟁력 및 경영성과와 직결된다. 소비자는 자동차의 ‘신뢰성’이 기대되는 시점에 자동차가 작동하지 않는다면 제품과 기업을 불신하게 되어 자동차 기업의 시장경쟁력은 저하될 수밖에 없다. 전 세계적으로

1997년 200여 건에 불과했던 연간 자동차 리콜 건수는 2017년 700건 이상으로 늘어나면서 자동차 결함 예방의 중요성이 증대되고 있다 (그림 1). 예로서 현대차그룹은 2020년 10월 세타2 GDI엔진의 품질비용으로 3조 3,600억 원의 충당금을 쌓기로 하였고 LG는 GM 전기차 배터리 리콜비용으로 1조 4,000억 원을 분담하기로 하였다.

또한, 제품수명주기에서 신뢰성 분석이 적용되는 시점이 빠르면 빠를수록 수익은 늘어난다. 잘 알려진 1:10:100의 법칙에 따라 제품 수명주기의 후반부(리콜)의 사후 조치보다는 전반부의 예방조치를 하는 것이 필요하다 (그림 2)[2]. 즉 초기 시장 출시 전 실험 데이터의 면밀한 분석은 자동차 기업의 경영성과가 직결된다.

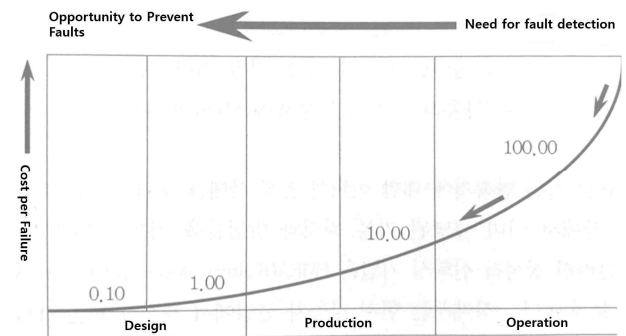


그림 2. 고장 비용과 제품 수명주기의 관계
 Fig. 2. The relationship between cost of failure and product life cycle

오늘날 최신 자동차의 개발에서 다양한 기능추가에 대한 요구, 높은 복잡성, 하드웨어/소프트웨어/센서 기술의 통합, 제품 및 개발 비용의 감소, 시장출시 주기의 감소 등으로 신뢰성은 더욱 중요한 요인으로 부각되고 있다. 자동차의 ‘친환경성’, ‘고효율’이 부각되면서 하이브리드 자동차(하이브리드차), 플러그인 하이브리드차(PHEV), 전기자동차(전기차, EV), 수소연료전지자동차(수소차, FCEV) 등이 빠르게 내연기관차를 대체할 것으로 전망되고 있다 (그림 3).

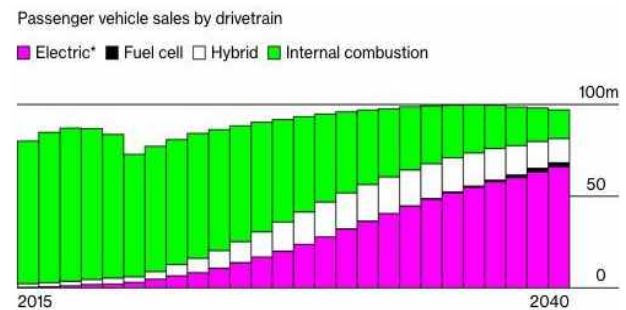


그림 3. 내연기관차와 친환경차의 판매대수 예측 (출처 : Green Car Reports)
 Fig. 3. Prediction of sales of internal combustion and eco-friendly vehicles (Source: Green Car Reports)

친환경차의 복잡성, 하드웨어/소프트웨어/센서 기술의 통합, 제품 및 개발 비용의 감소, 시장 출시 주기의 감소 등으로 신뢰성은 더욱 중요한 요인으로 부각되고 있다. 특히, 자율주행 기술의 도입으로 자동차의 ‘판단’이 개입되는 레벨 3부터 레벨 5까지의 신뢰성 성능평가를 위한 고도의 분석이 필요하다(그림 4). 왜냐하면, 자율주행 기능의 신뢰성이 운전자, 보행자의 안전, 생명과 직결되기 때문이다.

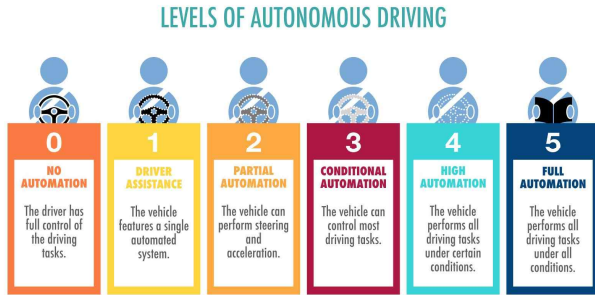


그림 4. 자율주행 단계 (출처: barco.com)
 Fig. 4. The stages of autonomous driving (source : barco.com)

그러나 현재 자동차 제조기업들의 주행데이터에 대한 분석은 전통적인 품질 개념의 통계 분석에 기반하여 고가의 통계 패키지를 활용하고 있으며 상시적인 분석이 아닌 일회성 보고서 작성으로 끝나는 경우가 많다. 실제 현장에서는 Control Chart를 이용한 정적인 산포도의 관리에 집중하고 있는 실정이다. 또한, 데이터 분석과 통계량 산출을 위하여 고가의 패키지를 사용하기 때문에 상당한 소프트웨어 라이선스 비용을 지불하고 이 패키지를 배우는데 드는 학습비용도 만만치 않다. 더 나아가 자동차 주행 품질관리의 결과 보고가 일회성 보고서로 끝나기 때문에 자동차 기업이 신뢰성 향상을 위한 상시적인 프로세스를 기업에 도입하고 정착시키는 데 어려움이 크다.

이에 따라 자동차 기업들은 자동차 주행시험평가 분석에서 기존의 전통적인 품질관리에서 ‘신뢰성 관리’로 빠르게 전환할 필요가 있다. 품질과 신뢰성은 표 1과 같이 큰 차이점이 존재한다. 가장 큰 차이는 시간에 대한 고려이다.

표 1. 품질과 신뢰성의 차이 (출처 : 한국신뢰성 협회)

Table 1. The difference between quality and reliability

	Traditional Quality Concept	Reliability
Time Perspective	Static	Dynamic
Focus	Present	Future (Prediction,
Improving Method	Narrow the Spread	Know the Liimits
Model	Normal Distribution for Product Characteristics	Exponential, Wiburll, lognormal Distribution
Main Interest	Product Characteristics	Product Life
Indicator	Defective Rate	Life span, failure rate, reliability

세계 자동차 산업은 친환경화-지능화-서비스화 등의 패러다임 변화가 가속되는 가운데 2030년 미래차 시장은 전기·수소차, 자율주행차, 이동서비스 산업이 주도할 것으로 전망된다. 현대자동차그룹은 2025년까지 전 세계에서 100만대의 전기차를 판매한다는 계획에 따라 브랜드와 회사별로 전동화 전환 전략을 구체화하고 있다. 일본의 자동차 업체들 역시 구미시장과 중국 시장의 변화를 염두에 두고 대응책 마련을 서두르고 있다. 한국을 비롯하여 미국, EU 등 각국 정부 역시 탈탄소화를 위한 친환경차 도입 계획을 공격적으로 발표하고 있다.

이와 같은 새로운 시대에 대비하여 자동차 제조기업들의 차량 주행 시험 평가의 대상과 방식의 근본적 혁신으로 4차 산업혁명 시대에 걸맞는 주행성능 시험 프로세스의 정립이 필요하다. 기존 자동차 주행 분석 방식의 문제점을 해결하기 위하여 주행데이터 분석에서 빅데이터 및 AI 기법을 활용한 신뢰성 분석 프로세스와 이를 구현하는 플랫폼을 개발하는 것이 필요하다. 전통적인 신뢰성 분석기법과 함께 최적의 머신러닝 알고리즘을 적용한 신뢰성 분석 모델을 개발하고 적용함으로써 분석력, 예측력, 결과제시 능력을 극대화하는 것이 필요하며 Python이 제공하는 오픈소스 패키지인 Pandas, Numpy, Reliability, Seaborn, Scikit Learn 등의 패키지를 활용해서 데이터를 가공함으로써 소프트웨어 라이선스 비용을 제로화하는 시도가 필요하다. 또한 일회성 신뢰성 분석이 아닌 상시적인 신뢰성 향상을 위한 혁신 프로세스를 자동차 제조기업이 도입할 수 있도록 신뢰성 분석 AI 플랫폼의 기능을 정의하고 구현하는 것이 중요하다 (그림 5)[3, 4].

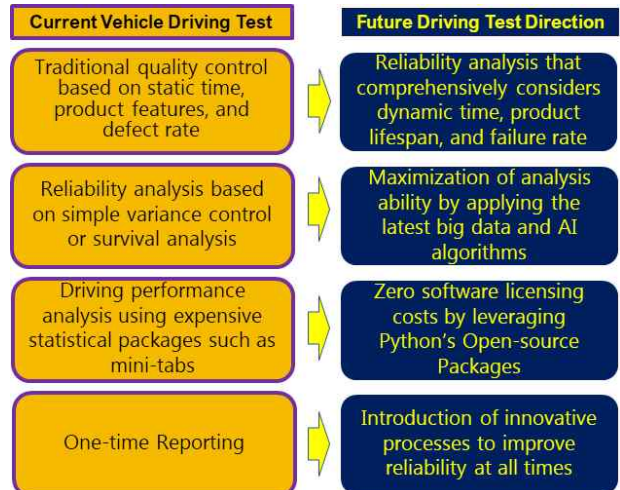


그림 5. 자동차 주행성능 평가의 변화 방향
 Fig. 5. The direction of improvement on vehicle driving performance evaluation

따라서 본 논문의 목적은 신뢰성 이론과 인공지능을 활용하는 주행 평가 프로세스를 정립하고 이로부터 시스템 기능과 아키텍처를 도출한 후 주행 빅데이터 플랫폼 프로토타입을 API (Application Programming Interface) 우선 (First)

개발방법론에 따라 개발하는 것을 목표로 한다.

주행 빅데이터 신뢰성 분석 AI 플랫폼은 자동차 제조업들의 차량주행 성능 평가 방식을 근본적으로 혁신하여 품질 높은 자동차 생산을 도모, 이는 소비자 혜택으로 이어질 수 있다. 자동차 주행성능 분석은 자동차 제조기업의 핵심적인 프로세스이고 분석의 수요가 매우 커 산업적 파급효과가 매우 큰 과제이다. 자동차 제조회사라면 당연히 주행성능 테스트를 수행하게 되며 이 과정에서 대량의 데이터가 발생하고 분석의 수요도 크며 분석 결과의 활용이 자동차 기업의 경쟁력과 직결된다고 할 수 있다.

II. 본 론

2-1 시스템 기능 도출과 아키텍처의 설계

자동차 주행 신뢰성 분석 시스템의 주요 기능을 도출하고 그 기능들을 구현하기 위한 타겟 아키텍처의 설계를 위하여 시스템개발 프로세스를 우선 정립하였다. 그리고 이를 기반으로 시스템의 기능을 도출하고 도출된 기능을 컴포넌트화한 후 도출된 컴포넌트를 연계하여 시스템 아키텍처를 설계하였다. 이 과정에서 시장 지향적 요구사항 수집과 상업성 높은 비즈니스 모델에 기반한 시스템을 기획하는데 초점을 맞추었다. 즉, 최종 사용자인 엔지니어의 니즈를 기반으로 시장 지향적인 시스템을 기획하고 분석, 설계하는 것을 큰 방향으로 잡았다.

1) AI 기반 자동차 주행 신뢰성 평가 프로세스의 정립

신뢰성 평가 프로세스는 차량 주행시험 빅데이터의 수집, 전처리, 분석 및 시각화의 전주기를 지원하는 활동들로 구성된다 (그림 6). 구체적으로 주행 테스트 데이터를 수집하여 탐색적 자료 분석과 전처리를 수행하는 활동을 진행한다. 이 데이터를 기반으로 전통적인 신뢰성 분석기법의 적용을 통한 인사이트를 도출한다. 다음으로 딥러닝, 머신러닝 분석기법을 활용하여 수명예측 인자를 발견하는 과정과 학습을 통해 수명예측 인공지능 모델을 도출한다. 그리고 수명예측 모델에 기반하여 효과적인 기술적, 관리적 대안을 도출한다. 최종적으로 신뢰성 분석 모델의 예측도 향상을 위한 데이터의 수집 개선방안을 구상하게 된다.

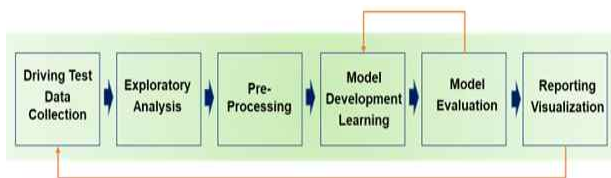


그림 6. AI 기술을 적용한 자동차 주행 평가 및 최적화 프로세스
 Fig. 6. Vehicle Driving Test and Optimization Process Using AI Technology

2) 시스템 기능의 도출

신뢰성 평가 프로세스를 바탕으로 시스템의 기능을 도출하였다 (표 2). 시스템의 기능들에 대한 구체적인 기술은 Unified Modeling Language(UML)의 유스케이스 (Use Case) 다이어그램 등을 통해 가능하다.

표 2. 차량 주행 빅데이터 신뢰성분석 AI 플랫폼의 기능
 Table 2. The Main Functionalities of Reliability Analysis AI Platform on Vehicle Driving Big Data

Classification	Main Functionality	Details
Exploratory Data Analysis	Data quality analysis	Measurement of data quality such as missing values and outliers, visualization of distribution by variable and calculation of characteristic statistics, correlation analysis and deduction of correlation coefficients using scatter plots and correlation heatmaps
Pre-Processing	Cleansing	Missing value processing, noise removal smoothing function, and contradiction removal function for effective application of algorithms and improvement of model results
	Reduction	Dimensional reduction through principal component analysis, numerical reduction through clustering
	Transformation	Discretization of Numerical Data, Log Transformation, On Hot Encoding, Feature Split
Individual Analysis	Statistical Analysis	Searching a list of 1,000 sample driving data and analyzing basic statistics for each variable using Pandas Profiling
	Visualization	Histogram analysis, Displot analysis, Driving location map display using Kakao Map, Log Time Series and Time Series, Heatmap, Pairplot analysis, Severity analysis for individual driving
Metadata Management	User Management	User registration, login, logout, password search, profile, activity log search
	Driving Data Management	Driving list inquiry, condition search, driving record creation/change/deletion, driving record comment inquiry/creation/change/delete
Reliability Analysis	Reliability Function	Derivation of reliability functions such as cooling water inlet and outlet temperature difference and control compliance
	Severity Analysis	Severity analysis and knocking analysis for specific variables
Comparative Analysis	Comparative Analysis	Boxplot, Histogram Comparison, comparative analysis of data sets, reliability monitoring by driving using features and variables
Correlation Analysis	Correlation Analysis	Correlation coefficient trend, Scatter Plot, Log Scale function
In-Depth Analysis	Diagnosis of malfunctions	Provides a diagnostic function that can evaluate and monitor vehicle performance degradation, damage, etc.

Discovery of influencers	Discovery of factors affecting vehicle failure through machine learning and evaluation of the importance of influence
Reliability Analysis	Failure evaluation and prediction function through application of traditional reliability analysis algorithm
Machine Learning Model Development	Failure prediction function through machine learning model learning and application

3) 시스템 아키텍처의 설계

플랫폼의 기능을 시스템 아키텍처를 설계하기 위하여 각 기능을 구현하는 시스템의 구성단위인 컴포넌트와 컴포넌트들을 계층적으로 모은 레이어(Layer)를 구성하였다. 레이어는 백엔드, 애플리케이션, 프론트엔드로 분류된다.

백엔드 레이어는 주행시험 데이터, 메타데이터, 모델베이스로 구성되며 파일 시스템과 데이터베이스 혼용 데이터 레이크의 설계를 통해 유연한 구조를 만드는데 초점을 두었다. 애플리케이션 레이어는 전처리 로직, 학습 로직, 분석 로직을 포함할 수 있도록 구조화하였다. 프론트엔드 레이어는 캐싱과 인메모리 데이터 처리와 함께 가벼운 클라이언트 렌더링 로직 적용을 통해 성능을 극대화하였고 논리적이고 미려한 사용자 인터페이스를 구현하여 사용자와 효과적으로 상호작용하는 것을 목표로 설정하였다.

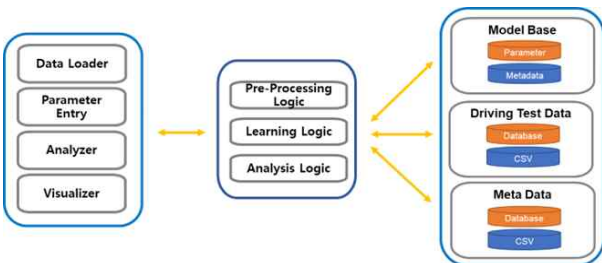


그림 7. 차량주행 빅데이터 신뢰성 분석 플랫폼의 시스템 아키텍처

Fig. 7. The System Architecture of Vehicle Driving Big Data Reliability Analysis Platform

2-2 자동차 주행 신뢰성 분석 시스템의 개발

본 시스템을 개발하는 과정에서 다양한 사용자 요구사항을 수용하면서 신속한 개발을 도모하고 유지보수성을 향상시키기 위하여 「API First」 기법을 도입하였다. API 우선 개발 방식은 프론트엔드의 요구사항과 백엔드 데이터베이스 설계 요구사항을 함께 수용하여 웹 프로토콜을 이용하여 실행되는 함수를 애플리케이션 레이어 내에서 우선 개발하고 API로 구현한 후 이를 기반으로 프론트엔드와 백엔드를 구현하는 방식을 의미한다.

1) 애플리케이션 레이어 REST API의 설계와 구현

클라이언트와 서버간의 상호 작용을 최적화하고 효율화하기 위하여 REST API를 이용하였다. REST API는 무상태(Serverless) 방식으로 동작되므로 임의의 순서로 리소스를 요청할 수 있고 다른 요청과 분리되어 동작된다. 또한 REST API를 통해 계층화된 시스템으로 서비스를 조직화하면서도 확장성 있게 서비스를 구성할 수 있다. 사용자들은 웹기반의 Client와 REST API를 통해 다양한 형태의 자원들을 정형화된 인증과 인가과정을 거쳐 체계적으로 접근하는 구조를 만들 수 있다.

표 3. REST API 목록

Table 3. A list of REST API

Classification	REST API	Description
User Management	/user/registration	User Registration
	/user/login	Log-In
	/user/logout	Log-Out
	/user/password	Password Finding
	/user/withdrawal	Withdrawal
Driving Record Management	/drive	Driving Record List-Up
		Creation of Driving Record
		Update of Driving Record
		Delete of Driving Record
Driving Analysis	/sample	Sample Data Presentation
	/variable	Variable Description
	/statistics	Basic Statistics and Histogram
	/map	Presenting Driving information on the map
	/timeseries	Time
	/severity	Severity Indicators calculation and presentation metadata
	/knocking	Knocking analysis data request

데이터 흐름 다이어그램 (Data Flow Diagram, DFD) 기법을 활용하여 자동차 데이터의 수집, 저장, 변환, 분석 과정을 설계한 후 이를 단위 플로우로 분석한 후 각각의 플로우를 REST (Representational State Transfer) API로 구현하였다. 데이터 흐름 다이어그램을 이용한 분석방법은 전처리, 분석 로직 등 복잡한 프로세스 중심의 프로그램을 구현하는 데 적합하다.

데이터 플로우 다이어그램을 바탕으로 REST API를 우선 정의한 후 백엔드와 프론트엔드를 구현하는 방식은 사용자 요구사항의 변경 등에 대해 신속하게 대응이 가능하며 프론트엔드와 백엔드의 독립성을 강화하여 유지 보수성을 향상시킬 수 있다.

REST API를 구축하기 위하여 파이썬 패키지 중 가장 빠른 성능을 자랑하는 FastAPI를 이용하였다. FastAPI는 Django,

Flask에 비해 높은 처리속도를 자랑하며 Node.js에 버금가는 성능을 보여 준다. 자주 활용되는 데이터에 대한 캐싱과 인메모리 데이터 처리를 통한 성능을 극대화하였다. 그림 8와 같이 REST API들은 Swagger UI를 통해 조회할 수있다.



그림 8. Swagger UI를 이용한 REST API 목록 조회
Fig. 8. The List of REST API through Swagger UI

파이썬의 Airflow 자동화 패키지를 이용하여 DFD 요소들을 구현하였다. 자동화 패키지는 Flow File, Processor, Connection, Flow Controller, Processor Group으로 구성되는데 이는 DFD의 설계 구성요소들과 잘 맵핑된다. 즉, 데이터 흐름은 FlowFile, Connection과 맵핑되고, 프로세스는 Processor와 Processor Group과 맵핑되어 쉽게 로직을 구현할 수 있다.

2) 백엔드 레이어 데이터베이스의 설계와 구현

데이터 플로우 다이어그램의 데이터 스토어를 데이터 레이크로 설계하는 방식으로 백엔드를 구현하였다. 데이터 레이크란 대규모의 다양한 원시 데이터셋을 저장하는 리포지토리를 의미하며 정형, 반정형, 비정형 데이터를 모두 포괄한다. 데이터스토어는 ‘정형’데이터 형태의 데이터베이스 이외에 ‘반정형’ 형태의 분석결과물, 이미지, 동영상 형태의 ‘비정형’ 분석결과물을 모두 포함해야 한다. 파일시스템과 데이터베이스를 혼용하는 데이터 레이크의 설계를 통해 이 모든 산출물을 구조화하고 저장과 질의의 성능을 최적화하였다. 구체적으로 데이터 레이크에서 관리하는 데이터는 사용자 인터페이스의 시각화 라이브러리들이 요구하는 형태로 저장되어 별도의 변환 절차 없이 빠르게 차트 렌더링과 그리드 생성이 가능한 형태로 구조화하고 파일 시스템에 저장하였다. 데이터베이스 내부의 데이터는 Query를 이용하여 변환 절차를 거친 후 사용자 인터페이스로 전달되나 데이터 레이크의 결과물은 변환 절차가 필요 없이 빠른 성능 실현이 가능하다. 다양한 전처리, 데이터 엔지니어링 로직 등을 적용하기 위해 필요한 원시데이터와 적용 후 산출되는 결과물을 json, csv, 리스트, 이미지, 동영상 형태로 저장할 수 있는 리포지토리 (파일시스템)

가 데이터레이크의 물리적인 구조이다.

데이터레이크의 구조화를 위해서 본 연구에서는 아마존 웹 서비스 상의 RDS (Relational Database Service)와 S3 (Simple Storage Service)를 활용하였다. RDS는 클라우드에서 간편하게 데이터베이스를 설치, 운영 및 확장할 수 있는 관리형 서비스 모음으로서 PostgreSQL 엔진을 사용하여 정형화된 데이터를 구조화하여 처리하였다. S3는 데이터 가용성, 보안 및 성능을 제공하는 객체 스토리지 서비스로서 인증, 인가에 대한 규칙을 적용하여 구성하였다. RDS와 S3를 AWS의 Glue와 같은 다양한 인터페이스 툴을 이용하여 유기적으로 통합하여 두 시스템 간에 데이터가 자유롭게 오갈 수 있도록 하였다.

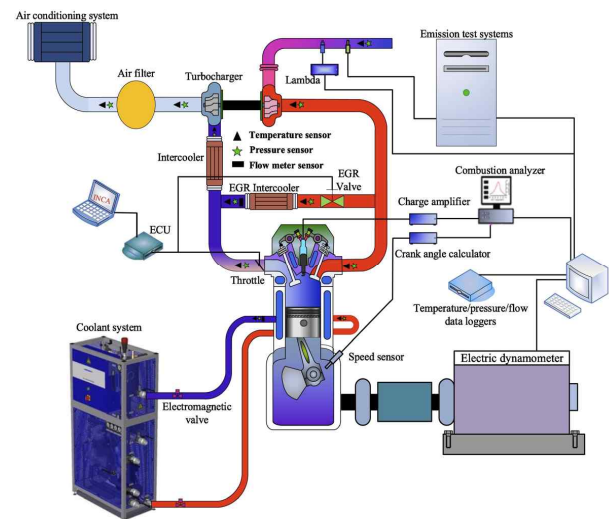


그림 9. 내연기관 자동차 내 센서의 부착 위치
Fig. 9. Location of sensor attachment in an internal combustion engine vehicle

또한 분석 결과물들은 다양한 차원에서 형태적, 내용적으로 분류된 후 메타데이터에 객체(Object)로 등록되는 절차를 거치게 된다. ‘메타데이터’란 데이터 구조에 관한 데이터로 사용자와 원시데이터, 분석 결과물들을 유형화하고 데이터의 출처, 소유자, 갱신시기, 디렉토리 등을 기록하는 별도의 데이터베이스이다. 앞서 언급된 주행 데이터베이스의 다양한 View들과 데이터레이크 내 산출물들은 객체(Object)로 정의하고 분류한 후 메타데이터에 등록한다. 객체는 개별주행 또는 분석의 입력물이나 결과물이 될 수 있다. 객체간 관계 (예: 원시데이터 셋과 분석 결과물)를 정의할 수 있고 객체의 물리적 존재 형태(query, view, json, csv, 이미지, 동영상 등)와 장소를 기록하며 객체를 만든 사용자, 활용한 사용자, 활용된 횟수, 갱신기록, 사용자의 코멘트 등도 기록된다. 또한 사용자 등록, 보안, 이력 관리 기능을 논리적으로 설계하고 구현한다. 사용자는 사용자 그룹에 할당되고 각 그룹에 대해서 객체(Object)에 대한 CRUD 접근 권한을 제어한다. 사용자가 로그인한 기록, 사용자가 객체를 만들고 할당하고 조회하고 삭제

제하고 갱신하고 코멘트한 기록도 저장된다.

데이터 레이크에는 자동차의 다양한 계측데이터와 특성 모수 그리고 분석지표들을 논리적으로 저장하기 위한 구조가 만들어진다. 자동차 주행 데이터는 자동차 각 부위에 센서를 부착하여 계속되는데 차속, 엔진토크, 변속단수, 배기온도, 토크, 모터 냉각 유온, 인버터, 배터리 온도 등 다양한 변수가 센서를 통해 수집된다(그림 9) [5].

수집된 원시데이터는 가공과정을 거쳐 신뢰성 관련지표, 가혹도 지표, 운전자의 주행습관을 특징화할 수 있는 주행 패턴 특성 패러미터, SOH (State of Health), 전비, 모터 출력 등 다양한 지표로 산출된다[6, 7].

표 4. 가혹도 지표
Table 4. Severity Indicator

Severity Feature	Feature Details
length	Length of time series
absolute Energy	Sum over the squared values
absolute_sum_of_changes	Sum over the absolute value of consecutive changes in the series x
mean_abs_change	Mean absolute change
cid_cc	An estimate for a time series complexity (A more complex time series has more peaks, valleys etc.)
count_above (t)	Percentage of values in x that are higher than t (threshold)
count_below (t)	The percentage of values in x that are lower than t
number of peaks	Calculates the number of peaks in the time series
ratio_beyond_2_sigma	Ratio of values that are more than 2 * std(x) (so r times sigma) away from the mean of x
approximate entropy	A technique used to quantify the amount of regularity and the unpredictability of fluctuations over time-series data
sample_entropy	A modification of approximate entropy (ApEn), used for assessing the complexity of physiological time-series signals approximate entropy
maximum	the highest value of the time series
minimum	the lowest value of the time series
absolute Maximum	the highest absolute value
crest factor	Peak(Absolute Maximum)/RMS
peak to peak	maximum-minimum

개체관계도를 이용하여 운영데이터 저장소와 스타스키마, 메타데이터를 설계함으로써 개발생산성과 유지보수성 향상을 도모하였다. 또한, 차량 주행 데이터베이스에 대한 다양한 View들을 효과적으로 도출할 수 있는 데이터웨어하우스 스타스키마를 구현하였다. CASE (Computer Aided Software Engineering) 도구를 활용하여 데이터베이스를 설계함으로써 사용자, 시스템 아키텍트, 개발자 간 의사소통을 향상시킬 수 있었다.

백엔드의 구현 과정에서 오픈 DBMS를 활용하여 구축함으로써 소프트웨어 라이선스 비용을 제로화하고 시스템의 유연

성과 확장성을 도모하였다. 오픈 DBMS 중 기술적으로 가장 앞서 있는 PostgreSQL Server Version 14를 적용하여 데이터베이스를 구축하였다. 시스템 응답속도 ‘2초 이내’를 목표로 최신의 백엔드 기술을 적용하였으며 파일시스템 (json, csv, list set)과 데이터베이스를 혼용하는 데이터 레이크의 설계를 통해 쿼리 성능을 최적화하였다.

3) 프론트엔드 레이어의 구현

프론트엔드의 구현을 위하여 컨텐츠에 최적화된 시각화 기술을 적용하고 이들을 효과적으로 연계하는 하이퍼미디어 설계 기술을 적용하였다. Histogram, Scatter plot, Box Plot, Displots, Heatmap 등 다양한 시각화 기술 중 컨텐츠를 가장 잘 표현할 수 있는 기법을 적용하였으며 차량 실주행 데이터베이스의 View들을 효과적으로 조합하고 연계하는 하이퍼미디어 기술을 적용하여 네비게이션 구조를 설계하였다. W2UI.js, uplot.js, D3.js 등 가벼운 자바스크립트 오픈 시각화 패키지를 적극 활용하여 클라이언트 렌더링 성능을 극대화하였다.

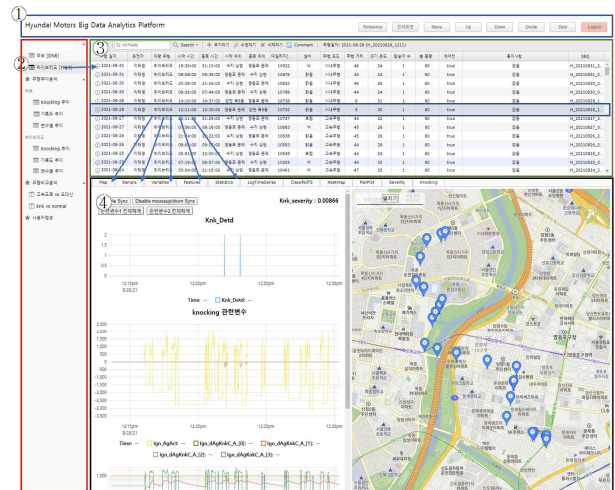


그림 10. 하이퍼미디어 기반의 사용자 인터페이스 구성
Fig. 10. Construction of User Interface based on Hypermedia

인터페이스 구성 방법은 사용자 편의성, 하자보수 편의성 강화를 위한 UI 표준을 수립하여 적용하였고 미숙련자도 쉽게 사용 가능할 수 있도록 직관적인 사용자 인터페이스를 구현하였다[8]. 사용자 인터페이스와 네비게이션은 그림 10과 같이 하이퍼미디어 개념에 따라 논리적이며, 직관적이고 미려하게 구성하여 사용자는 사고의 흐름에 따라 논리적으로 네비게이션을 수행할 수 있다. ① 화면은 헤더 화면으로 사용자는 인터페이스에 대한 설정과 로그아웃을 수행하며 변경되지 않고 항상 상단에 위치한다. ② 화면은 주행 차량, 주행추이 기능, 주행비교 기능 등 메타데이터 내 객체들의 Collection, Class, View Type 등에 대한 Anchor Point를 트리 형태로 구조화하였다. ③ 화면은 ② 화면의 각 Anchor 포인트에 대한 인스턴스들을 그리드 형태로 표현하였다. ④ 화면은 ③ 화

면에서 선택된 각 인스턴스에 대한 시각화 구조물을 담은 컨테이너로서 차트, 그리드, 매쉬업, HTML 등 다양한 시각화 결과물을 탭으로 구분하여 제시할 수 있다.

2-3 개발 시스템

개발된 시스템의 기능은 크게 개별주행분석, 주행 추이분석, 주행 상관분석, 주행 비교 분석, 심층주행 분석으로 구조화되어 있다.

1) 개별주행 분석

수집된 개별 주행데이터는 정제, 집계를 통한 축소, 주성분 분석을 통한 차원축소, 로그변환, 이산화 등의 전처리과정을 거친다. 그리고 개별주행의 시계열 패턴을 확인하기 위한 로그 단위 오실로스코프, 변수들의 빈도 패턴 분석을 위한 히스토그램, 변수별 통계량, 주행 경로 파악을 위한 지도 매쉬업등이 제공된다.

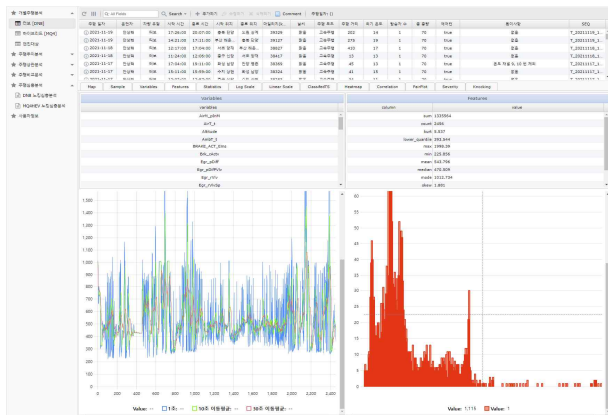


그림 11. 개별주행 분석을 위한 Feature 추출
Fig. 11. Feature Extraction for Individual Driving Analysis

2) 주행추이 분석



그림 12. 변수의 장기 추이
Fig. 12. Long Term Time Series for a Variable

장기간에 걸친 주행시험 결과 추이를 한눈에 모니터링하고 차량의 성능 저하, 손상 발생 등을 탐지할 수 있는 기능을 제공한다. 성능 저하와 손상 발생 심도의 평가를 위한 진단함수를 개발해서 적용할 수 있다. 또한 가혹도 산출을 통해 각 주행마다 다른 주행 패턴과 도로특성을 표준화할 수 있다.

3) 주행상관 분석

성능 저하, 손상 발생을 탐지하고 확인한 후에는 고장에 영향을 미치는 영향요인을 밝혀내는 것이 필요하다. 히트맵(Heatmap)을 통해 엔지니어는 수십, 수백 개에 이르는 주행 변수 간 상관 강도와 계수를 한눈에 확인할 수 있고 Pair Plot 을 통해 문제해결의 단서를 찾을 수 있다. 고장 유형별로 관련 변수를 조합하여 화면을 별도로 구성하는 것도 가능하다.

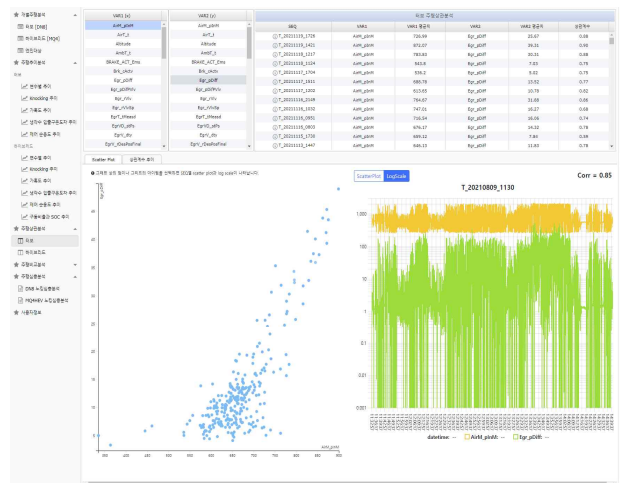


그림 13. 변수 간 상관분석
Fig. 13. Correlation Analysis of Two Variables

4) 주행 비교 분석

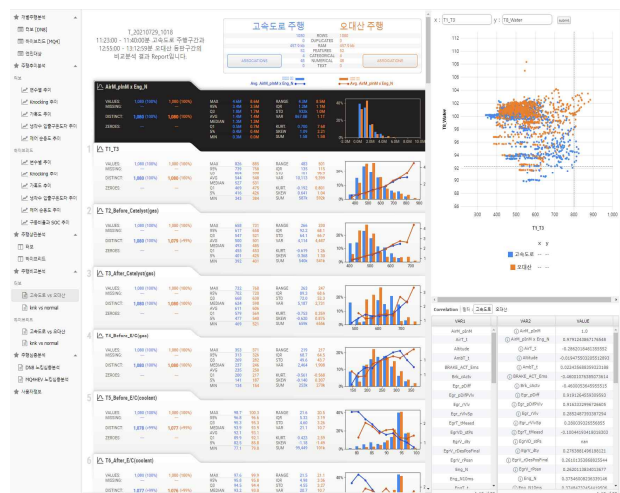


그림 14. 주행 간 비교분석
Fig. 14. Comparison Analysis among Driving Data

주행 간, 주행 내 구간 간, 또는 특정 조건을 만족하는 주행 요소를 샘플링하여 상호 비교하게 되면 성능 저하의 원인과 기능개선의 기회를 파악할 수 있다. 또한 하드웨어적으로, 소프트웨어적으로 차량 성능을 개선한 후 개선 전후의 주행시험 결과를 비교함으로써 그 효과를 확인할 수 있다.

5) 심층 주행 분석

차량의 고장 진단, 성능 개선에는 많은 변수들이 동시에 개입되기 때문에 매우 복잡한 양상을 띠게 된다. 본 시스템은 머신러닝, 딥러닝을 통해 차량의 고장에 영향을 주는 인자를 발견하고 중요도를 평가한 후 대안을 수립하고 검증하는 기능을 제공한다.

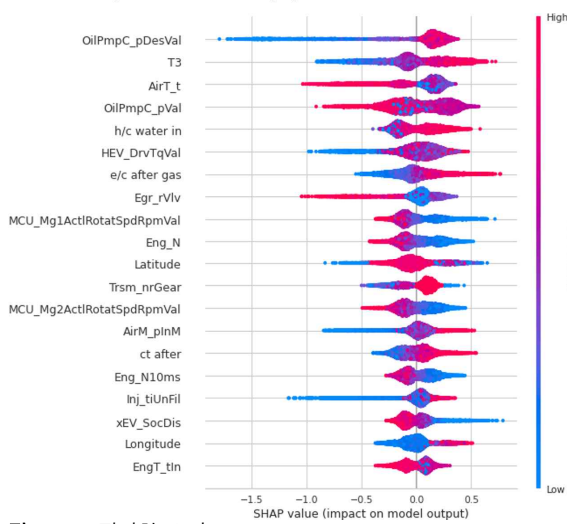


그림 15. 고장영향 요인 SHAP plot
Fig. 15. SHAP plot of the factors for the disorder

III. 결 론

현재 자동차 기업들의 주행데이터 분석 서비스는 전통적인 품질 개념의 통계 분석에 의존하고 있고 일회성 보고서 작성으로 끝나는 경우가 많으며 공정관리도 (Control Chart)를 이용한 정적인 산포도의 관리에 집중하고 있는 실정이다. 또한, 데이터 분석과 통계량 산출을 위하여 고가의 패키지를 사용하기 때문에 상당한 소프트웨어 라이선스 비용을 지불하고 이 패키지를 배우는데 드는 학습비용도 만만치 않다.

이러한 문제점을 해결하기 위하여 본 연구는 하이브리드, 전기차, 수소차 등 친환경차의 부상으로 변화하는 새로운 자동차 산업 생태계에서 자동차 제조기업과 부품업체들의 주행 성능 평가방식을 근본적으로 혁신할 수 있는 시험 프로세스의 정립과 AI 플랫폼의 개발이라는 목표 아래 신뢰성 분석에 기반의 최신의 품질 관리 프로세스를 자동차 성능 테스트에 도입하기 위한 최적의 프로세스와 AI 기술을 이용한 플랫폼을 구현하였다.

또한 정부는 「2030 미래자동차 산업 발전전략」 추진을 통한 “2030년 미래차 경쟁력 1등 국가로의 도약”을 비전으로 제시하고, 이러한 비전을 달성하기 위해 구체적인 목표를 설정한 바 있다. 이와 같은 정부의 정책 방향을 실현하고 기업들의 친환경차 개발 노력을 가속화 시키기 위해서는 빅데이터에 기반한 주행 성능 분석 프로세스와 플랫폼의 개발이 필요하다.

개발된 플랫폼은 내연기관, 하이브리드, 전기자동차 등 다양한 차종을 대상으로 실시되는 엔진대상, 스피드웨이 주행시험, 실주행 시험 과정에서 발생하는 빅데이터의 수집, 전처리, 분석, 시각화 전 주기를 지원한다.

이 플랫폼이 지향하는 동적 신뢰성 분석, 최신 빅데이터/AI 알고리즘의 적용, 소프트웨어 라이선스 비용의 제로화, 상시적인 신뢰성 향상 프로세스의 도입을 통해 자동차/부품 기업들은 진정한 품질지향기업으로 도약할 수 있다. 신뢰성 분석 AI 플랫폼은 자동차 제조 기업들의 차량주행 성능 평가 방식을 근본적으로 혁신하여 품질 높은 자동차 생산을 도모하고 이는 소비자 혜택으로 이어질 수 있다.

본 연구는 자동차의 고장 유형으로서 노킹에 초점을 맞추고 머신러닝을 이용한 심층 주행분석을 수행하였다. 따라서 본 연구는 한 가지 고장 유형에 대한 분석에 그치고 있어 다양한 고장 분석에 적용할 수 있는 일반화된 플랫폼의 구조와 기능을 제시하지 못하는 한계를 지니고 있다. 향후 다양한 고장 유형에 대한 주행 분석 결과를 바탕으로 엔지니어들의 신뢰성 분석 업무를 효과적으로 지원할 수 있는 시스템의 구조와 기능을 제시하는 것이 필요하다.

감사의 글

본 연구는 2022년도 중소벤처기업부(중소기업기술정보진흥원)에서 시행한 창업성장기술개발사업의 지원에 의하여 이루어진 연구로서, 관계부처에 감사드립니다.

참고문헌

- [1] P. A. Tobias, D. C. Trindade, *Applied Reliability*, 3rd ed. New York, NY: A Chapman & Hall Book, 2011.
- [2] B. Bertsche, *Reliability in Automotive and Mechanical Engineering*, Berlin, Heidelberg: Springer-Verlag, 2008.
- [3] S. Lee, B. Lee, J. Chung, S. Choi, D. Kim, “A Study on the Real Road Operating Characteristics of Vehicle Components through Element-trip Analysis,” in *Proceeding of Korean Society of Automotive Engineers Spring 2021*, pp. 60-61, 2021.
- [4] H. J. Jung, M. S. Noh, K. W. Rho, “Analysis Study of Passenger Vehicle Fuel Economy according to Different Driving Resistances,” in *Proceeding of Korean Society of*

Automotive Engineers Spring 2020, pp. 150-154, 2020.

- [5] K. W. Rho, H. J. Jung, C. H. Sim, "A Study on Causality between Vehicle Cumulative Mileage and Fuel Economy Using Statistical Analysis Method," *Transactions of the Korean Society of Automotive Engineers*, Vol. 29, No. 6, pp. 589-595, 2021.
<https://doi.org/10.7467/KSAE.2021.29.6.589>
- [6] J. W. Lee, C. B. Chung, I. C. Choi, "Severity Factors affecting Tire Wear," *Ploymer(Korea)*, Vol. 29, No. 1, pp. 48-53, 2005.
- [7] S. H. Kim, S. H. Kim, B. Yoon, "Deep Learning-Based Vehicle Anomaly Detection by Combining Vehicle Sensor Data," *Journal of the Korea Academia-Industrial cooperation Society*, Vol. 22, No. 3, pp. 20-29, 2021.
<https://doi.org/10.5762/KAIS.2021.22.3.20>
- [8] J. M. Lee, "A Study on Dashboard Convergence Design for Data Visualization," *The Korean Society of Science & Art*, Vol. 38, No. 5, pp. 423-435, 2020.
<http://doi.org/10.17548/ksaf.2020.12.30.423>



김종호 (Jong-Ho Kim)

1994년 : KAIST 경영정책학과 (공학사)
1996년 : KAIST 경영정보공학과 (공학석사)
2003년 : KAIST 경영공학과 (공학박사-빅데이터)

1996년~2003년: 비트컴퓨터
2003년~2006년: 삼성SDS
2006년~2008년: 가톨릭대학교 연구조교수
2008년~2011년: 현대경제연구원
2011년~현 재: 경성대학교 경영학과 부교수, 주식회사 하이퍼로직 대표
※ 관심분야 : 빅데이터, 인공지능, 시스템 분석 및 설계



박화규 (Hwa-Kyu Park)

1990년 : 미 캘리포니아주립대 산업 및 경영 석사 (경영학석사)
1993년 : 미 오클라호마주립대 산업 및 경영학 박사 수료 (경영학 박사 수료)
1998년 : KAIST 경영정보-의료경영 (경영학박사)

1993년~2004년: KIST 한국과학기술연구원 연구원 SERI ETRI 한국전자통신연구원 보건의료정보부문 선임연구원
2005년~현 재: 순천향대학교 의료과학대학 교수
※ 관심분야 : 의료정보, 의료경영, 의료통계