Journal of Convergence for Information Technology
Vol. 11. No. 6, pp. 270-281, 2021
e-ISSN 2586-4440
DOI : https://doi.org/10.22156/CS4SMB.2021.11.06.270

# 기계학습을 이용한 블록체인 기반의 보험사기 예측 모델 연구

**이용주**
**충북대학교 연구원**

# A Study on the Blockchain-Based Insurance Fraud Prediction Model Using Machine Learning

YongJoo Lee
Researcher, Division of Software, Chungbuk National University

**요 약** 정보기술의 발달로 보험사기의 규모는 매년 급증하고 있고, 그 방법도 공모 형태로 조직화되고 고도화되고 있다. 이를 예측하고 검출하기 위한 다양한 형태의 예측모델이 연구되고 있지만 보험관련 정보는 매우 민감하여 공유와 접근에 위험이 높고 법적인 혹은 기술적인 제약이 많다. 이 논문에서는 최근 4차 산업 혁명의 등장으로 가장 각광받는 기술 중 하나인 블록체인을 기반으로 한 기계학습 보험사기 예측모델을 제안한다. 블록체인 기술을 활용하여 안전하고 신뢰받는 보험청구 정보 공유시스템을 실현하고, 보다 효율적이고 정확한 사기예측을 위하여 사회관계분석이론을 적용하여 각 관계에 가중치를 부여하고 기계학습 사기 예측패턴을 4단계로 나누어 제안하였다. 사기 가능성이 높은 보험청구건은 보다 앞선 단계에서 높은 예측 율로 검출되는 효과를 가지며 가능성이 낮은 청구 건은 사후에 참고하여 관리할 수 있도록 차등 적용하였다. 제안하는 모델의 중요 매커니즘은 이더리움(Ethereum) 로컬 네트워크를 구성하여 검증 하였고, 향후 보다 정교한 성능평가가 요구된다.

**주제어 :** 블록체인, 기계학습, 보험사기, 이더리움, 지도학습

**Abstract** With the development of information technology, the size of insurance fraud is increasing rapidly every year, and the method is being organized and advanced in conspiracy. Although various forms of prediction models are being studied to predict and detect this, insurance-related information is highly sensitive, which poses a high risk of sharing and access and has many legal or technical constraints. In this paper, we propose a machine learning insurance fraud prediction model based on blockchain, one of the most popular technologies with the recent advent of the Fourth Industrial Revolution. We utilize blockchain technology to realize a safe and trusted insurance information sharing system, apply the theory of social relationship analysis for more efficient and accurate fraud prediction, and propose machine learning fraud prediction patterns in four stages. Claims with high probability of fraud have the effect of being detected at a higher prediction rate at an earlier stage, and claims with low probability are applied differentially for post-reference management. The core mechanism of the proposed model has been verified by constructing an Ethereum local network, requiring more sophisticated performance evaluations in the future.

**Key Words :** Blockchain, Machine Learning, Insurance Fraud, Ethereum, Supervised Learning

*This work was supported by Institute for information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No. 2019-0-00708, Integrated Development Environment for Autonomic IoT Applications based on Neuromorphic Architecture).

*Corresponding Author : YongJoo Lee(yjlee3363@naver.com)

Received April 29, 2021
Revised May 23, 2021
Accepted June 20, 2021
Published June 28, 2021

## 1. 서론

4차 산업 혁명의 등장과 더불어 블록체인 기술은 핀테크라는 금융계의 혁신으로 떠오르면서 주목받기 시작하였고, 암호화폐 기능을 넘어 무한한 잠재력과 가능성을 세계적으로 인정받으며 세계경제포럼(World Economic Forum)에서 미래를 바꿀 10대 기술 등으로 선정되며 블록체인 관련 투자 또한 급증하고 있다. 블록체인은 분산 네트워크의 복수 노드에 정보를 저장하고 공유하는 분산원장 기술을 기반으로 구현된 응용기술이다. 기존의 중앙 집중형 네트워크에서 발생하는 중앙관리자의 위험, 노드 장애의 위험 등을 제거하고 새로운 형태의 탈중앙화 네트워크(Decentralized Network)에 기반하므로 중앙 관리자 없이 당사자만으로도 실시간으로 정보를 공유할 수 있고 동시에 당사자로 인한 변조 등의 위험을 제거할 수 있게 된다[1]. 안전한 데이터의 저장 및 교환을 위해 해시알고리즘, 디지털서명, 공개키 알고리즘, 합의알고리즘 등의 기술이 사용된다. 최초의 블록체인은 비트코인이며, 다양한 비즈니스 모델을 타겟으로 개발된 이더리움(Ethereum)은 당사자 간 미리 계약된 내용의 조건이 성사되었을 때 자동으로 성사되는 기술을 구현한 스마트 컨트랙트를 제공하며 금융, 무역, 사물인터넷 등의 다양한 응용서비스 등에 활용될 수 있다. 블록체인 기술을 이용한 다양한 스타트업이 소개되면서 스마트 컨트랙트 기술 또한 주목받고 이를 기반으로 한 거래, 투표, 보험, 의료, 부동산 등 고도의 보안성의 요구되는 분야에 다양하게 응용 및 연구되고 있다. 스마트 컨트랙트를 제공하는 이더리움 플랫폼은 자체에서 채굴된 암호화폐 이더(Ether)를 제공하며 다양한 비즈니스 모델에서 트랜잭션을 전송하고 저장하기 위해 거래된다[2].

보험사기는 보험 상품 가입자가 공급자를 기망하여 부당한 이익을 편취하는 것을 말하며 그 금액은 매년 급격히 증가하고 있다. 보험사기로 2017년에 적발된 금액은 7천억 원을 넘었고 이는 2007년 금액의 4배에 달한다. 2018년에는 상반기에만 4천억 원을 넘어서 매년 증가추세를 보이고 있다. 또한 적발된 보험사기 중 손해보험에서 발생한 사기가 90프로에 달하며 이중에서 자동차보험의 보험사기 비율이 가장 높다. 이렇게 높은 수치에도 불구하고 모든 보험사기가 적발되는 것은 아니므로 실제 그 건수와 금액은 더 높을 것으로 예상한다. 보험사기의 양만 증가하는 것이 아니라 정보기술의 발달로 보험사기 방법은 점점 더 치밀하고 고도화 되고 있으며 단순 범죄에서 벗어나 전문가가 주도하는 공모형의 범죄가 늘고 있다. 그럼에도 불구하고 보험 관련 정보는 매우 민감한 개인정보 등이 포함되어 있기 때문에 공유하기가 힘들고 다른 형사 사건 등에 비해 해결의지가 낮아 전문 수사 인력의 배치에도 어려움을 겪는다. 따라서 매우 비슷한 사기 수법으로 여러 회사의 보험금을 부당청구하고도 적발되지 않는 사례가 발생하며 이는 결국 다른 계약자의 보험료 인상으로 이어져 모두에게 피해를 입히게 된다[3].

보험사기를 방지하고 이미 발생한 건에 대한 적발을 높이기 위해서는 관련 기관 간의 정보 공유가 필수이나 여러 보호법안의 제약으로 쉽지 않다. 한국신용정보원에서 운영하는 보험사고정보시스템이 있으나 신용정보법에 정보를 집적하므로 정보제공에 동의한 정보만을 사용할 수 있어 유용한 정보 공유 등에 한계가 있고, 실제 보험사기 적발 및 조사에 서로 협력하여야 하는 여러 단체 간 정보 공유에 대한 표준화된 틀이 없어 작업에 어려움이 많다. 보험사기를 사전에 예방하고 발생한 청구 건에 대한 사기를 적시에 적발하기 위해서는 보험회사의 보험사고 관련 정보에 대한 용이한 접근성이 반드시 필요하며 여러 보험회사와 관련 기관과의 공유를 위한 표준화된 폼이 필요하다. 또한 이러한 공유를 통해 어떠한 신용정보나 개인정보의 노출이 발생하지 않도록 보안측면에서 완벽한 방안이 준비되어 있어야 할 것이다. 그러나 기존의 중앙 집중 시스템에서는 중앙관리자를 통해 매우 민감한 정보를 공유하기 때문에 이에 대한 위험이 높고 시스템 해킹 등으로 인한 정보 유출의 위험성 또한 심각하다.

이 논문에서는 보안성과 신뢰도가 높은 블록체인을 기반으로 하는 기계학습을 이용한 보험사기 예측 모델을 제안하고자 한다. 이어지는 2장에서는 보험사기 관련 연구 및 기계학습을 이용한 보험사기 관련 연구 등을 소개하고 3장에서는 블록체인 기반의 보험사기 예측모델, 4장에서는 기계학습을 이용한 예측 모델에 대해 다루고 5장에서 이에 대한 평가 및 실험 내용을 정리한다.

## 2. 관련연구

### 2.1 금융이상거래 추출을 위한 관련연구

비대면 채널을 통한 금융 관련 사기가 사회적 이슈

로 급부상 하면서 이에 대한 관심과 사회적인 대책 등이 제안되어 왔다. 금융 관련사기를 막기 위한 대부분의 연구는 대응방안과 현 실태분석, 그리고 법적인 제도 마련에 초점이 맞추어져 있고 공학적 관점에서는 금융거래 이상 탐지에 초점이 맞추어져 있다[4]. 금융거래 이상패턴을 감지하기 위해 블랙리스트 기반 추출과 이상거래 패턴을 기반으로 한 탐지를 기본적으로 시행한다. 보다 높은 탐지 율로 예측하기 위해 금융환경에서 제공하는 통계적 기법, 빅 데이터 분석기법, 매체정보를 중심으로 한 탐지 기법 등을 기반으로 연구되고 있다. 블랙리스트 기반 탐지는 정탐 율은 높지만 예측 빈도가 낮고 이상 패턴 기반은 다양한 기법의 사기에 대해 유용하게 사용될 수 있지만 블랙리스트 기법에 비해 오탐율이 높고 많은 데이터 축적을 필요로 한다. 또한 동일 사고 발생 시 실제 데이터로 등록되기 전까지 비슷한 수법의 사기가 발생할 수 있어 패턴 등록을 위한 절차의 신속성도 요구된다. 정의석은 현재 이상탐지 기법의 단점을 보완하고 보다 고도화된 사고 등에 대응하기 위하여 인공지능 기계학습 알고리즘을 도입하여 실시간 탐지가 되는 이상금융거래탐지 모델을 연구하였다[5]. 2017년에 A은행에서 일어난 사고 등의 데이터를 축적하여 딥러닝 기반 지도학습 방식의 CNN(Convolutional Neural Network) 알고리즘을 선택하여 탐지모델을 시연하였다. 지도학습 기반의 탐지모델을 구현하기 위하여 해당 은행의 사고데이터를 수집하고 데이터 전처리를 시행하여 이상패턴 추출 모델을 생성하였다. 본 연구는 기존의 데이터를 기반으로 이상패턴을 추출하는 기존의 지도학습 모델을 금융거래 이상패턴에 적용한 예이다. 이상거래의 과거 데이터에 소개되지 않는 신종 사기 수법이라면 해당 사기는 실시간으로 검출하지 못하게 되며 이상금융거래로 의심 받았다 하더라도 실제 사기 건으로 분류되어 정보가 공유되기 까지 많은 시간이 소요되어 많은 피해를 초래하게 되는 한계가 있다.

### 2.2 보험사기 예방을 위한 관련연구

보험사기를 검출하기 위해서는 피보험자의 정보를 기반으로 한 고의성 추출이 핵심 기술이다. 다양한 형태의 보험사기가 존재하지만 동시에 비슷하게 여러 보험에 가입하여 과도한 금액을 챙기는 사기 형태 등은 매우 보편화된 사기 수법에 속한다[6]. 금융감독원에서는 보험사기인지시스템(Insurance Fraud Analysis System: IFAS)를 개발하여 국민과 보험회사를 통해 입력된 정보를 기반으로 보험사기를 예측하고 수사를 의뢰하는 시스템으로 활용하고 있다. Fig. 1은 이러한 유관기관과 보험회사, 수사기관과의 공조를 보여주고 있다.

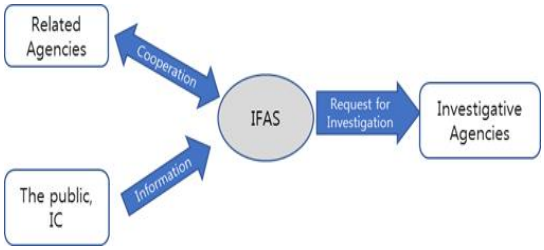


**Fig. 1. Architecture of IFAS**

IFAS는 보험사기를 청구 당시에 적발하고자 하는 의도도 있지만 사후에 의심자를 적발하여 향후 다른 사기를 예방하기 위한 목적도 있다. 한국신용정보원에서 제공하는 “보험사기다잡아” 시스템도 이와 같은 목적으로 운용되고 있다. 생명보험사와 손해보험사 공제사 등이 보험정보를 공유하는 방식으로 보험계약정보 및 청구정보를 실시간으로 공유해야 하는 매우 민감하고 어려운 문제를 해결하여야 한다[7].

### 2.3 이상패턴 예측을 위한 기계학습 관련연구

다양한 분야에서 이상패턴 추출을 위해 기계학습(Machine Learning) 모델이 사용되고 있다. 이는 레이블이 있는 지도학습(Supervised Learning)과 비지도학습(Unsupervised Learning)으로 구분한다. 지도학습에서 대표적인 예측모델은 회귀(Regression) 모델, 분류(Classification) 모델, 의사결정트리(Decision Tree) 등이 있다. 의사결정 트리(Decision Tree)는 수치 자료나 범주 자료 등에 가장 널리 쓰이는 지도학습 모델이다[8]. 대규모 데이터 셋에서도 잘 동작하는 모델로서 의료보험사기 검출, 일상생활에서 이상패턴 등의 검출 등의 연구에도 널리 사용되고 있다. 서포트 벡터 머신(Support Vector Machine)은 최근 가장 널리 쓰이는 분류 및 예측 기법이다. 벡터 공간에 위치한 학습 데이터가 선형으로 분리 가능한 선형 SVM과 분리하기 어려운 비선형 SVM으로 나뉜다[9]. 서포트벡터를 기준으로 결정경계가 1보다 크면 정상으로 분류하고 -1보다 작으면 이상으로 분류하는 방법이며 마진을 최대화하기 위해 라그랑주 함수를 도입하여 최적해를 구하는 문

제로 해석할 수 있고 터널 함수(Kernel function)을 이용하여 정확도를 높힌다[10, 11]. 로지스틱 회귀 모델은 지도 학습의 대표적인 예측 모델로 다양한 응용분야에서 이상패턴, 정상범위를 벗어나는 행위, 이벤트 등의 관찰을 식별하는데 용이하게 사용된다. 로지스틱 회귀모델을 보험사기 판별에 이용하기 위해서 부당청구(Fraud)와 정당청구(Non-Fraud)의 두 가지 값을 갖는 이진 종속변수를 정의하고 결과에 영향을 미치는 요인들 $(x_{1,}x_{2,,,}x_{p})$이 입력변수로 사용되고 이를 수식으로 나타내면 식(1)과 같다.

$$y=\beta_0+\beta_1x_1+\ldots+\beta_px_p+\varepsilon, \quad (1)$$
$$y=\begin{cases}0, Fraud\\1, Non-Fraud\end{cases}$$

결과는 정상 범위를 벗어날 수 있어 일반적인 가정을 충족시키지 못하는 문제점이 발생할 수 있고 이를 위해 연속성을 만족하는 증가함수인 연결함수를 통하여 모델링 하는 것이 로지스틱회귀모형이다. 이를 나타내면 식(2)와 같다.

$$\Pr[y=1|x|=q(\beta_0+\beta_1x_1+\ldots+\beta_px_p) \quad (2)$$

연결함수의 형태에 따라 여러 형태로 나타날 수 있으며 회귀계수를 추정하기 위해서 최대가능도 추정 방법을 사용한다. 이상패턴을 검출하기 위한 기계학습 연구는 다양하게 진행되고 있다.

### 2.4 블록체인 기술을 이용한 사기탐지 연구

블록체인은 생성되는 거래를 블록단위로 체인처럼 연결하여 저장하는 기술로서 한번 저장되면 수정이나 삭제가 어렵고 타임스탬프가 포함되어 시간의 순서대로 연결되므로 관련 사건의 추적이 용이하다. 반면 누구에게나 공개되어 있어 정보 침해의 위험이 높아 이에 대한 수준 높은 대책이 요구된다. 이러한 블록체인 기술은 당사자만으로도 시간의 경과에 따른 명확한 소스가 제공된다는 점에서 높은 보안성을 요구하는 다양한 분야에 연구되고 있다[12].

고자영 외는 블록체인 거래에서 발생하는 사기, 불법거래 등의 이상탐지를 위하여 기계학습 알고리즘을 이용하였다. 비트코인 거래내역 중 도난 및 해킹 등에 사용된 트랜잭션을 데이터로 확보하고 이상거래 탐지에 활용하였다[13]. 특히 과표본 추출기법으로 가장 대표적인 SMOTE(Synthetic Minority Oversampling Technique)을 도입하여 훈련데이터 셋을 제공하여 과적합 문제를 해결하였고 로지스틱회귀 방법과 SVM을 이용하여 이상거래 분류를 시행하였다. 이 연구는 블록체인을 통해 수행된 거래 중 이상이 감지되는 거래를 추출하는 것에 목적이 있고, 블록체인 플랫폼을 기반으로 사기탐지에 활용하는 것은 아니어서 보험사기 예측에 직접적인 활용은 어렵다. 비트코인 이상거래를 검출하기 위해 비정상 트래픽을 추출하기 위한 특징들을 연구들도 진행되고 있고 특히 장기진은 블록체인 기술을 이용한 혁신적인 금융서비스에 관하여 연구하였다. 본 연구에서 블록체인은 단순한 핀테크 개념으로서만이 아니라 이상거래를 방지하기 위한 정보보안, 결제보안 등을 연구하였으나 본 연구도 거래 자체에 대한 보안향상에 초점이 맞추어져 있는 한계가 있다[14].

최근 보험시장에서 보험계약과 관련된 정보를 관리하는 방안으로 블록체인을 이용하는 연구가 진행되고 있는데 이는 공개키 방식과 디지털 서명등의 암호 알고리즘을 통하여 서로 안전하게 거래하고 기록이 실시간으로 보존되며 변조가 어려운 특징을 보험계약에 이용하고자 하는 것이다[15, 16]. 현재 보험 등에 블록체인을 적용하는 사례는 이러한 거래의 효율성 특면과 자동으로 당사자 간에 거래가 성사되는 스마트 컨트랙트의 활용성을 이용한 보험계약의 분야가 두드러진다[17-19]. 보험 분야에서 가치사슬은 이러한 보험상품 등의 개발이나 고객에 대한 영업, 고객 자산의 운용 및 보험금 청구와 고객 정보 관리 등으로 구성되며 암호화폐를 통한 보험금 지급과 비용지불 등도 가능한 연구범위에 해당된다. 최근 보험회사는 컨소시엄을 구성하거나 블록체인을 접목한 새로운 비즈니스 모델 등이 소개되고 있다[20, 21]. 향후 단순한 보험거래를 위해 스마트 컨트랙트를 활용하는 수준에 그치는 것이 아니라 기존의 보험사기를 검출하는 방법 등을 블록체인을 통해 실현하여 보안성과 안전성을 제공하면서 동시에 실시간으로 최신 사기 패턴까지 검출해 낼 수 있는 모델 등에 대한 연구가 필요하다[22, 23].

이 논문에서 제안하는 보험 사기탐지 기법에 블록체인 기술을 이용하는 이유는 기존의 관련연구들과 같이 블록체인 거래의 사기 탐지가 목적이 아니다. 현재 실

생활에서 이루어지는 보험사기 탐지의 문제점으로 지적되는 안전한 보험 청구 정보공유의 문제 등을 해결하기 위하여 블록체인 플랫폼을 이용하여 예측모델을 설계하고자 하는 것이다. 안전하게 정보보호를 하면서 동시에 인증 받은 사용자만 공유할 수 있도록 하는데 그 목적이 있다.

## 3. 블록체인 기반의 보험사기 예측모델

### 3.1 보험사기 예측모델의 개요

기존의 보험사기 예측 모델은 기존의 사기 정보를 기반으로 사기 가능성을 추출하는 것에 초점이 맞추어져 있으므로 새로운 형태의 사기 등에 대처하기 어려운 문제점들이 있다. 이러한 단점을 극복하기 위해서는 사기정보 뿐만이 아니라 현재 청구되는 보험사고 내용을 서로 공유하고 보험사를 바꾸어 가며 사기를 반복하는 반복사기 형태까지 예측할 수 있어야 한다. 그러나 사기 패턴 정보가 아닌 현재 청구되는 보험 청구 내용은 매우 민감하고 중요한 정보를 포함하고 있어 이를 공개하고 공유하는 것은 어려운 일이다. 이를 해결하기 위한 방법으로 블록체인 기반의 보험사기예측모델(Blockchain-based Insurance Fraud Prediction Model(BIFPL)을 제안한다. 다수의 보험회사는 고객들의 정보를 블록체인을 통해 등록하고 차량 사고로 인한 보험청구가 발생할 시 해당 정보를 블록체인에 등록하고 기계학습 패턴예측을 통하여 사기 가능성을 확인하게 된다.

### 3.2 BIFPL의 요구사항

이 논문에서 제안하는 블록체인 기반의 보험사기 예측 모델을 실현하기 위해서 다음과 같은 요구사항을 필요로 한다. 첫 번째로 실시간 청구 내용을 공유할 수 있는 플랫폼이 제공되어야 하는데 이때 거래되는 중요 정보는 안전하게 보호 되어야 하며 다른 목적으로 악용될 수 없어야 한다. 제안하는 방식에서는 보험 청구에 포함되는 내용을 보호하기 위하여 암호기술을 이용한다. 해당 내용을 그대로 공유하는 것이 아니라 단방향 함수인 해시 알고리즘을 이용하여 문제를 해결하고자 한다. 같은 정보를 갖고 보험사기 예측을 요청하는 보험회사만이 해당 내용을 파악할 수 있기 때문에 외부에 공개되지 않을 뿐만 아니라 다른 보험회사들도 해당 내용을 확인할 수는 없게 되어 악용 가능성을 제거할 수 있다. 또한 블록체인의 특성상 중앙 관리자 등에 의한 정보 저장이 없어 중요 정보는 입력을 변환한 최초 등록자만이 알고 있으므로 중간자에 의한 공격 등으로부터 안전하다.

두 번째로 보험회사를 가장한 악의적인 사용자로부터 악용당하지 않기 위하여 인증 받은 보험회사만 예측정보를 받을 수 있어야 한다. 따라서 신뢰할 수 있을만한 인증방법이 필요하다. 일반적인 블록체인 거래는 사용자가 블록체인으로 트랜잭션을 발생시키는 구조이지만, 이 경우에는 블록체인 내에서 외부 서버로 트랜잭션을 발생시켜야만 해당 보험회사의 인증을 요청할 수 있게 되는데 이는 현재 오라클라이즈 기술을 이용하여 해결할 수 있다. 마지막으로 실시간으로 공유되는 정보들을 저장하여 기계학습 모듈에 활용하기 위해서는 보다 효율적인 특징추출 전처리블록을 위한 인터페이스가 존재하여야 하며, 처리 결과를 앞당길 수 있는 효율적인 매커니즘이 필요하다. 제안하는 방식에서는 이를 해결하기 위해 기계학습 처리 기능을 가중치를 참고하여 네 단계로 나누어 보험사기를 예측하는 모델을 제안한다. 가중치가 높아 사기 가능성이 높은 사건에 대하여는 보다 빨리 대응할 수 있도록 하였다.

### 3.3 블록체인기반의 보험사기예측모델 구조

Fig. 2는 블록체인 기반의 기계학습을 이용한 보험사기 예측모델을 위한 BIFPL의 네트워크 구조도를 나타낸다. 블록체인을 중심으로 연동하는 보험회사 인증서버(ICAS, Insurance Company Authentication Server)와 머신러닝 블록(ML, Machine Learning)이 연결되어 있고, 다양한 보험회사(IC, Insurance Company)들이 블록체인에 연결되어 있다.

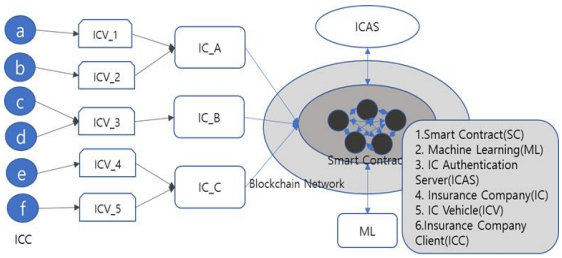


Fig. 2. Architecture of BIFPM

스마트 컨트랙트는 이더리움으로 운영되는 블록체인 플랫폼을 의미하며 블록체인은 복수의 노드에 구현된 컨트랙트가 배포되어 운영되고 있다. 스마트 컨트랙트의 이용자인 보험회사(IC)는 블록체인 계정(account)을 가지고 있으며 이더리움의 암호화폐인 이더(Ether)를 보유하고 지불하며 한 쌍의 공개키$(p_{ic_i}, q_{ic_i})$를 보유하며 식(3, 4)와 같다.

$$ic \in IC \quad (3)$$

$$\forall\ ic_i \ni ic_i(p_{ic_i}, q_{ic_i}) \quad (4)$$

보험회사는 여러 보험가입자(ICC, Insurance Company Client)를 고객으로 관리하고 있으며 각 고객은 보험청구의 주체로써 고유한 아이디를 보유한다. 고객의 보험청구 내용은 다른 보험회사들과 정보를 공유하여야 하는데 고객의 개인정보나 신원 등이 외부에 공개되어서는 안 되므로 이를 보호할 수단이 필요하다. 제안하는 매커니즘에서는 정보보호를 위한 보안성과 효율성을 모두 만족시키기 위하여 고객 고유번호 (ex: 주민등록번호)를 해싱하여 생성되는 32바이트에서 고유성이 부여된 20바이트만을 사용한다. 보험가입자는 보험청구 시 사기여부를 판단하기 위한 기본정보 (직업(j), 나이(a), 성별(s), 전과(cr), 기존 청구기록(bh) 등)이 포함되며 식 (5, 6)와 같다.

$$icc \in ICC \quad (5)$$

$$\forall\ icc_j \ni icc_j(j_j, a_j, s_j, cr_j, bh_j \ldots) \quad (6)$$

보험회사가 스마트 컨트랙트에 관련 트랜잭션을 등록하면 해당 내용은 모든 노드에 복사되고 업데이트 된다. 보험 상품이 가입되고 사고로 인한 보험청구가 되면 사기 가능성이 있는지 사기예측을 요청하게 된다. 가입자 차량(ICV, IC Vehicle)은 실생활에서 자동차보험을 가입한 차량을 의미한다. 보험가입자는 차량을 기반으로 보험 상품에 가입한 실제 차량이용자를 의미하며, 차량에 사고를 당한 피해자나 가해자와는 별도의 의미이다.

### 3.4 보험사기 예측모델 매커니즘 정의

보험회사는 보험사기 예측을 요청하기 위하여 첫 번째로 합당한 인증을 거쳐야 한다. 그림 3은 BIFPM의 전체 절차를 보여준다. 합당한 보험회사로 인증 받은 후 보험사기 정보를 등록하고 이후 해당 정보의 사기예측을 요청할 수 있다. 보험사기 예측은 보다 짧은 시간에 사기 가능성이 높은 사건들을 검출하기 위하여 그림 3과 같이 네 단계로 나누어 시행한다.

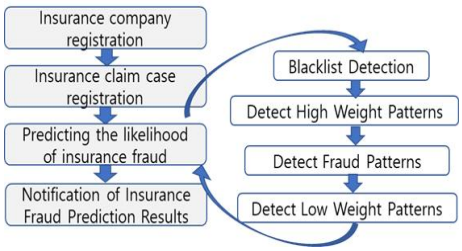


**Fig. 3. Procedures of BIFPM**

#### 3.4.1 보험회사 등록 단계

보험청구 내용은 외부에 공개되어서는 안 되는 정보이며, 공개되지 않고 예측 결과만을 공유한다 하여도 외부인이 이용할 수는 없는 매우 높은 보안성을 요구하는 정보이다. 따라서 해당 정보를 공유하고 사기 예측에 사용할 수 있는 보험회사는 인증을 마친 정상적인 보험회사로 제한한다. 인증을 마친 보험회사는 자신의 계정으로 보험청구 내용을 공유하고 사기예측을 요청할 수 있게 된다. 보험회사는 스마트 컨트랙트 프로그램의 사용자가 되며 계좌를 보유하고 등록 및 정보 공유를 위해 트랜잭션을 발생할 때마다 이더를 사용료로 지불한다. 보험회사는 오프라인 상에서 정상적인 보험회사 등록을 마친 회사이며 인증 서버를 통해 해당 요청을 하고 자신의 계정을 등록하는 절차를 거치게 된다.

#### 3.4.2 보험사기 예측 단계

보험사기 예측을 보다 효율적으로 처리하기 위하여 보험사기 가능성이 높은 단계부터 시행된다. 즉 보험사기 가능성이 높은 청구 건은 보다 이른 단계에서 높은 단계의 사기예측 결과를 통보받게 된다. 알고리즘 1은 스마트 컨트랙트에 보험청구 내용을 등록한 보험회사가 사기예측 결과를 요청하는 매커니즘을 단계별로 상세히 설명한 것이다. 보험청구 내용을 등록하기에 앞서 중요 내용은 단방향 함수인 해시코드로 변환하여 구조체 변수에 담아 전달한다.

---

*Algorithm 1 : Request_Fraud_Prediction*

---

*Input: hid, IRinfo, HWinfo, LWinfo*
*Output: result*
***Begin***
*1. detectBliacklist(hid) with value*
*: return result;*
*2. if result != veryhigh;*
*then detectHweightpattern(HWinfo) with value*
*: return result;*
*3. if result != high;*
*then requestIFP(IRinfo)with value*
*: return result;*
*4. if result != high;*
*then detectLweightpattern(LWinfo) with value*
*: return result;*
***end***

---

사기예측 결과는 Low/VeryLow 혹은 High/VeryHigh으로 분류된다. 각 단계에서 일정 유사도 이상이 예측될 경우 사기로 의심되고 의심되는 정도에 따라 다른 레벨의 결과를 통보한다. 사기예측 가능성이 높을수록 보다 이른 단계에서 결과가 통보된다.

## 4. 기계학습 보험사기 예측모델

### 4.1 보험사기 관련 개체의 SNA 분석

기계학습을 위한 사기패턴 추출을 위해서는 기존에 축척된 방대한 양의 학습데이터와 연산 시간을 필요로 한다. 그러나 블록체인을 이용한 스마트 컨트랙트에서는 요청한 결과를 시간 내에 전송해야하는 제약사항이 발생하므로 적은 데이터로 연산 시간을 줄이기 위한 검출 방법이 필요하다. 그럼에도 불구하고 보험사고의 유형은 점점 더 복잡하게 조직화되고 지능화되어 가고 있어 어느 하나의 패턴으로 사기를 단정하기 어려워지고 있다. 또한 단순 사기에서 다양한 형태의 공모 형으로 발전하고 있어 이에 대한 감지가 더욱 어려워지고 있다. 따라서 제안하는 방법에서는 짧은 시간에 보다 사기 검출 정확도가 높은 단계부터 차례로 시도할 수 있는 방법을 찾고자 하였다. 보험사기를 예측할 수 있는 정보들의 관계를 도출하고 가능성의 경중을 분석하여 보다 빠르게 사기 위험이 높은 정보들을 도출하기 위한 방법으로 SNA(Social Network Analysis)로 관계 도출이론을 반영하였다. 사회적 현상을 노드(개체)와 링크(관계)로 정의되는 네트워크로 구성하여 서로 관계된 구성원을 군집화하고 구조적 속성을 보다 명확하게 규명하기 위한 것이다[24, 25]. SNA 모델을 이용하여 네트워크의 유사성을 계산하는 과정에서 가중치 부여를 이용하여 유사성 지표를 생성한다. 가중치가 없는 경우에 Jaccard Simillarity(JC)를 이용하고 가중치가 있는 링크에는 Cosine Similarity를 이용한다. JC는 두 집합이 포함하는 데이터의 유사성을 비교하기 위해 기계학습 블록에서 보험사기 유형의 사기패턴 검증에도 활용된다[26].

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{7}$$

두 개의 집합(A, B) 사이에 유사도인 자카드 지수는 0에서 1 사이의 값을 가지며 위의 식(7)과 같이 정의된다. 다차원 양수공간에서 유사도를 측정하기 위한 cosine similarity는 식(8)와 같이 나타낸다. 다차원 링크구조를 갖고 있기 때문에 가중치를 갖고 있는 두 노드의 유사성을 측정하는 방법으로 사용한다.

$$\cos(\theta) = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \tag{8}$$

가중치가 없는 노드의 유사성과 가중치가 있는 노드의 유사성을 검출하는 방법을 적용하여 기계학습 패턴 추출에 활용하고 효율적인 패턴 검출 모델에 적용하고자 한다.

#### 4.1.1 가중치 검출을 위한 관계 도출

차량 보험청구 발생 시에 관련된 정보들은 사람, 차량, 병원, 사고패턴, 기타 정보들이 존재한다. 다양한 방법으로 보험사기를 공모할 수 있기 때문에 다각도에서

사기를 예측하고자 첫 번째로 사람과 관련된 노드를 정의하고 그들 간의 매트릭스를 구성하여 관계를 도출하고 가중치를 정의하고자 한다. 사람과 관련된 노드는 보험계약자, 보험청구자, 피해자, 가해자, 설계자, 기존 사기 전과자 등이 있다. 보험 계약자는 보험 청구자가 되며, 이때 보험 청구자는 피해자 일수도 있고 가해자 일수도 있으며, 피해자는 단독 피해자 혹은 공동 피해자가 될 수 있다. 그림4는 사람 노드들 간의 관계를 보여주는 네트워크이다.

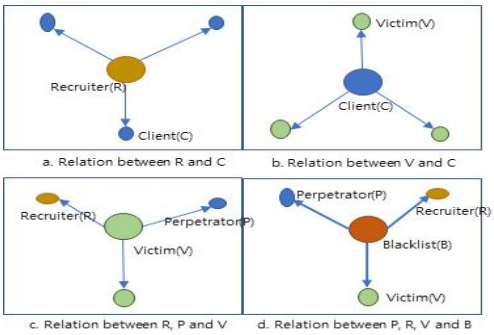


Fig. 4. Network between human nodes

Fig. 4에서 a는 설계사와 보험 계약자와의 관계를 나타낸 것이다. 한명의 설계사가 여러 명의 계약자를 가입시키고 다양한 수법으로 보험금 청구를 공모할 수 있다. 이는 매우 높은 가중치로 보험사기를 예측할 수 있다. b는 가해자와 피해자의 관계를 나타낸 것이다. 가해자가 여러 피해자와 관계를 가진다거나 동일 피해자와 다수의 사고가 발생하였다면 이도 역시 높은 가중치로 사기를 예측할 수 있다. c는 한명의 피해자가 다른 피해자, 설계사, 가해자와의 관계를 나타낸 것이다. 상대를 바꿔가며 비슷한 수법으로 보험금을 청구할 수 있기 때문에 높은 가중치로 보험사고를 예측할 수 있게 된다. d는 보험사고 청구에 관련된 사람 노드가 기존의 보험사기 전과 혹은 관련 의심이 있던 자로 분류된 블랙리스트와 관계를 보여주며 이 경우 매우 높은 사기 가능성을 갖게 된다. 이와 같이 사람 노드와의 관계를 갖는 링크는 높은 가중치를 갖고 사기 가능성을 검출하게 된다.

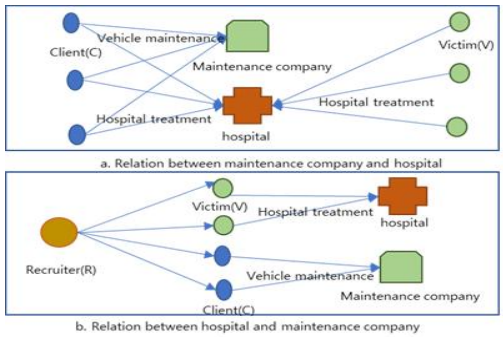


Fig. 5. Network between human nodes & companies

Fig. 5에서 a는 가해자가 특정 정비업체와 복수의 관계를 갖게 될 때 높은 가중치로 공모를 의심해 볼 수 있는 관계를 가지며 특정 피해자자 특정 병원에서 치료를 받게 될 때의 보험사기 가능성에 대한 관계를 설명한 것이다. b는 이러한 가해자와 피해자가 특정 설계사를 통해 보험에 가입하고 정비업체나 병원을 소개한 뒤 사기에 공모하는 관계를 보여주는 것이다. 사람 노드와 업체 및 병원과의 밀접한 관계가 매트릭스에서 보여 졌을 때 보험사기 예측결과가 뚜렷해진다.


Fig. 6. Type of Identity objects

Fig. 6은 보험사고 청구 시 갖게 되는 노드의 대표적인 종류를 보여주고 있다. 사람 개체로 설정된 노드나 차량 정비업체, 병원의 관계는 높은 가중치로 패턴을 검출하게 되며, 그 외 사람 노드가 갖게 되는 기타 정보인 소득, 직업, 신용등급 등은 낮은 가중치로 보험사기를 예측하게 된다.

### 4.2 보험사고 예측 상태 전이

SNA를 이용한 노드들 간의 관계를 도출한 것은 사기 가능성 예측을 하는데 필요한 시간 및 자원을 줄이고 보다 효율적으로 결과를 얻기 위한 것이다.

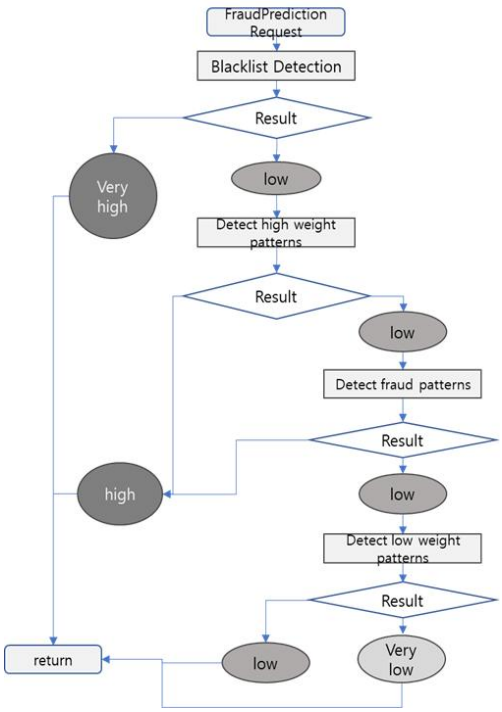


Fig. 7. Status Transition Diagram

높은 가중치를 갖는 관계에서 먼저 사기예측 검출 결과를 도출하고 차례로 저 가중치 예측을 시도하여 보다 짧은 시간에 정확성이 높은 검출 결과를 생산하고자 그림 7과 같은 상태전이를 이용하여 결과를 도출한다. 그림 7은 가장 높은 가중치로 사기를 예측할 수 있는 블랙리스트 비교 검출을 우선 실시하고, 이후 고가중치 검출 패턴을 실시 한 후, 사기패턴과 저 가중치 패턴을 검출하는 다이어그램을 보여주고 있다. 상위 단계에서 사기의심 결과가 출력된다면 매우 높은 수치로 예측되며, 하위 단계로 갈수록 높음과 낮음 매우 낮음의 상태로 이동하게 된다. 즉 사기가능성이 높으면 보다 빨리 검출할 수 있는 특징이 있으며 저 가중치에서는 검출 된다하여도 낮은 수치로 사기가 의심되는 차이가 있다. 즉 사기 전과자가 비슷한 수법으로 보험 청구를 한다면 사기의 가능성이 높지만, 소득이 낮고 무직인 사람이 보험 청구를 하였다 하더라도 사기의 가능성이 높은 것은 아니라는 뜻이며, 이는 보다 세심한 보험심사가 필요하다는 정도의 경고를 의미하며, 낮은 수치의 사기 가능성은 추후에 사기 검출을 위한 자료로 활용될 수 있다.

### 4.3 기계학습을 이용한 사기패턴 검출

차량 사고 발생 시, 사고 관련자나 사고처리에 관련된 정보 외에 사고의 내용도 중요한 가치를 지닌다. 비슷한 수법으로 사람을 바꿔가며 수차례 보험 청구를 하는 사례는 과거에도 많았으며, 수법 중 일부를 바꿔가며, 시기와 장소를 바꿔가며 반복적으로 시도하는 사기를 막기 위해서는 사기로 의심할 수 있을 만한 패턴 검출 방법이 필요하다. 과거에 사기로 판명 난 패턴 이외에도 신종 사기 수법이 있을 수 있으며 약간의 변형된 혹은 여러 수법이 혼합된 패턴 또한 검출할 수 있어야 한다. 사고가 발생하면, 사고와 관련된 사고 유형, 사고 장소, 사고 시간대, 사고 원인 등이 등록될 수 있다. 특정 사기는 인적이 드믄 오솔길에서만 가능하고, 특정사기는 여러 명이 공모하여야만 되는 교차로 등에서만 일어날 수도 있고, 특정 사기는 겨울 철 혹은 비오는 날 등에만 일어날 수 있기 때문이다. 이러한 유사패턴을 검출하기 위하여 청구사건 공유 시 사고 내용을 블록체인을 통해 제출하면 이에 대한 특성을 검출(Feature Engineering)하여 그룹 간 유사성을 계산하게 된다.

#### 4.3.1 그룹 간 유사 패턴 검출

기계학습 모델 중 기존의 다수의 사기패턴을 학습하여 이와 유사한 패턴을 검출하는 지도학습인 신경망 등을 이용하여 많은 관련연구가 진행되어 왔다. 그러나 기존의 사기 패턴을 학습하여 유사패턴을 검출하는 경우에는 신종 사기 및 공모 사기 등에 대한 예측이 어렵기 때문에 현재 공유되고 있는 보험청구 내용의 사고 패턴 등과 유사성을 비교하여야 한다. 따라서 학습할 데이터가 없이 현재 데이터를 기반으로 하는 비지도 학습을 통한 텍스트 분류를 통하여 데이터를 나누고 유사도를 검출하는 클러스터링 방식에 기반 하여 검출하고자 한다.

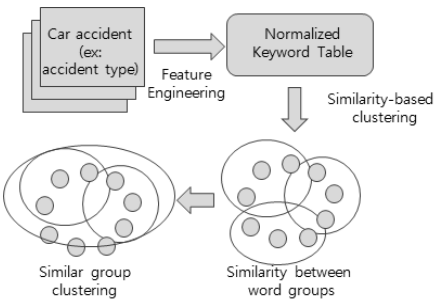


Fig. 8. Similar word group clustering

Fig. 8은 차량 사고 발생 시 공유되는 정보의 키워드를 추출하여 이에 대한 유사성을 기반으로 소그룹으로 클러스터링하고 클러스터링 된 그룹 간 유사도를 검출하는 단계를 보여주고 있다. 유사도는 유사한 단어로 군집된 소규모 그룹간의 유사도로 측정되어야 한다. 이를 위해 여러 문장을 단어의 집합으로 만든 뒤 집합을 통해 유사도를 측정하는 방법인 자카드 유사도 방식인 식(7)에 기반 한다.

## 5. 실험 및 검증

### 5.1 보안성 평가

보험청구 관련 내용은 극히 민감한 개인정보를 포함하고 있어 외부에 노출되거나 공유될 수 없는 정보이다. 이러한 문제를 어떻게 극복하는지에 따라 이 연구의 성과가 달려있다고 해도 과언이 아니다. 따라서 이 논문에서 제안된 보험사기 예측을 위해 공개되는 보험청구 내용을 보호하기 위한 프라이버시에 대해 평가한다. 해시 알고리즘은 단방향 함수로 입력이 다르면 출력이 다르고 입력의 1비트만 바뀌어도 출력이 완전히 바뀌어 입력 값을 예측하기 어려운 특징을 가지고 있다. 공개되는 사람 노드의 아이디는 해시코드로 변환하여 중요 정보가 포함된 20바이트를 사용한다. 타 단어와 구별될 수 있도록 고유성을 지니며 동시에 완벽하게 일치하는 ID에 대한 유사도는 추출할 수 있다. 사람 노드의 아이디뿐만 아니라 보험청구 내용 또한 중요 키워드를 추출하여 해시코드로 변환한 후 20바이트를 사용하는 방법으로 고유함, 효율성, 프라이버시를 모두 충족시킬 수 있도록 하였다. 현재 블록체인에서 사용하는 해시 알고리즘은 keccak 256비트 이며 이는 매우 높은 보안성과 계산 속도를 보여주고 있다. Table 1은 이에 대한 비교 자료이다.

**Table 1. Hasing Algorithm**

| Hash Function | Output Length | Security | Arithmetic speed |
|---|---|---|---|
| Md5 | 128bit | Low | Fast |
| Repemd1 | 160bit | Medium | Medium |
| SHA1 | 160bit | Medium | Medium |
| SHA256 | 256bit | High | High |
| SHA3 | 256bit | High | High |
| Keccak-256 | 256bit | High | High |

중요 아이디를 포함하는 키워드는 해싱되어 코드가 공유되기 사용되기 때문에 블록체인을 통해 그 정보를 해킹하거나 습득하여도 정확하게 같은 정보의 고객을 보유하는 보험회사 외에는 의미 없는 코드일 뿐이므로 개인정보 침해의 위험에서 벗어난다.

### 5.2 스마트 컨트랙트 기반의 네트워크 구축

보험사기 예측을 위한 모델의 블록체인 네트워크를 구축하여 구현 가능성을 시험하고 중요 알고리즘에 대한 실행 성능을 평가하였다. 시뮬레이션 환경은 solidity 버전은 “>=0.4.34 < 0.6.0”으로 이더리움 프라이빗 네트워크에서 실행하였으며 노드 정보는 Ethereum platform:“Geth/ v1.7.3-stable-4bb3c89d/ lunux-amd64/go1.9.3”/GANACHE/Window10/ Hardware:Intel CoreTMi7- 4790 CPU 3.6GHZ에서 수행하였다. 보험청구인과 보험회사는 모의클라이언트를 위하여 react와 nodeJS 모듈을 이용하여 간략히 구현하였으며 스마트 컨트랙트의 키는 METAMASK를 이용하여 관리하였고 시뮬레이션 및 배포를 위하여 트러플(Truffle)을 함께 사용하였다.

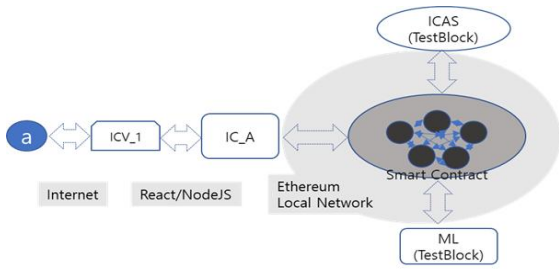


**Fig. 9. Environment of the simulation**

Fig. 9의 네트워크를 실제 구현 시 오라클라이즈로 구현되는 보험회사 인증 서버와의 연동과 기계학습 블록간의 연동은 테스트 블록으로 진행하였다. 이와 같은 환경을 실제 구현하여 Fig. 10과 같이 실행하였다.

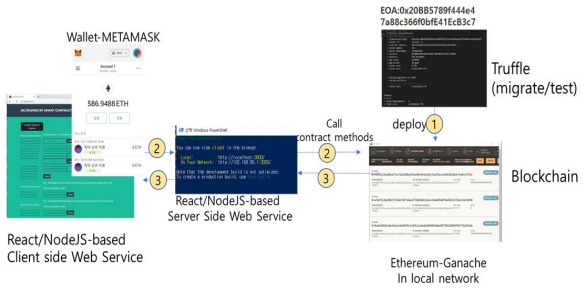


**Fig. 10. Simulated Network**

Fig. 10은 이더리움 로컬 네트워크에서 시뮬레이션 하는 전체 구조도를 보여준다. 보험회사(IC)는 전자지갑(METAMASK) 계정을 보유한 블록체인 이용자이며 이더(ether)를 보유하고 있다.

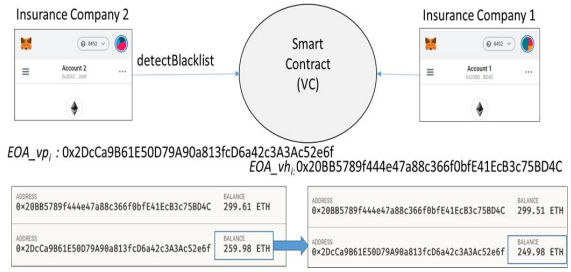


**Fig. 11. Ether changes after service delivery**

Fig. 11은 이더리움 로컬 네트워크에서 두 개의 보험회사(IC1, IC2)로 테스트 하였고 블랙리스트 검출을 요청하고 스마트 컨트랙트에서 요청한 서비스요금 10 이더를 지불한 이더의 변화를 보여주고 있다.

## 6. 결론

보험사기 규모와 금액이 매년 급증하고 있지만 이를 예측하고 검출하기 위한 시스템에는 여러 법적인 혹은 기술적인 제약사항들로 발전에 어려움이 많다. 사기로 판명나지 않은 신종 사기수법 등을 검출하기 위해서 블록체인을 통하여 보험청구 내용을 공유하였고, 이를 통해 개인정보를 안전 보호하는 동시에 신뢰받는 공유 시스템을 실현하였다. 또한 보험사기예측을 보다 정확하고 효율적으로 검출하기 위하여 공개되는 정보의 관계를 파악하여 가중치를 부여하고 이를 4단계로 나누어 기계학습을 이용한 예측 시스템을 제안하였다. 해당 시스템을 이더리움 기반의 로컬 네트워크에서 구현하여 모의테스트 하였다. 향후 테스트 블록으로 진행한 기계학습 기능과 보험인증 서버와의 연동도 실제로 구현하여 연동에 필요한 성능적인 점검도 필요하다.

## REFERENCES

[1] Y. David. (2017). Corporate governance and blockchains. *Review of Finance, 21(1),* 7-31. DOI : 10.1093/rof/rfw074

[2] S. S. Jeong. (2018). *Fintech Market Recent Trends and Implications.* Information Technology Promotion Center(Online). http://www.itfind.

[3] D. H. Park & D. J. Ryu. (2019). Information asymmetry in insurance market through blockchain technology: Focusing on the sharing of medical information and the prevention of insurance fraud. *Journal of the korean securities society, 48(4),* 417-447. DOI : 10.26845/KJFS.2019.08.48.4.417

[4] Y. H. Kim & K. J. Jang. (2021). Comparative analysis of predictive performance of ultrafine dust(PM2.5) based on deep learning algorithm, *Convergence Society for SMB, 11(3),* 7-13.

[5] Y. S. Jeong & J. I. Im. (2019). An Anomalous Intelligence(AI) detection Model Study of Telecommunications Finalcial Fraud Incidents, *KIISC, 29(1),* 149-164.

[6] K. D. Kim. (2017). The impact of blockchain technology on the insurance industry. *KIRI Report (Focus), 425,* 2-9.

[7] M. A. Lee. (2017). *Health Insurance Fraud Detection Using Data Mining Techniques.* Diss. Graduate School of Seoul National University, Seoul.

[8] X. Rui & D. Wunsch. (2005). Survey of clustering algorithms. *IEEE Transactions on neural networks 16(3),* 645-678. DOI : 10.1109/TNN.2005.845141

[9] C. Corinna & V. Vapnik. (1995). Support-vector networks. *Machine learning, 20(3),* 273-297. DOI : 10.1007/BF00994018

[10] P. Yi et al. (2006). Application of clustering methods to health insurance fraud detection. *In 2006 International Conference on Service Systems and Service Management* (Vol. 1, pp. 116-120). IEEE. DOI : 10.1109/ICSSSM.2006.320598

[11] P. Yi et al. (2007). Application of classification methods to individual disability income

insurance fraud detection. *In International Conference on Computational Science* (pp. 852-858). Springer, Berlin, Heidelberg.

[12] Y. S. Ko & H. S. Choi. (2017). Application and the business paradigm Changes and How to use them. *Korean Society of Science and Arts Convergence, 27,* 13-29.

[13] J. Y. Ko & S. J. Bae. (2019). A Study on the Detection Method of Blockchain Abnormalities Based on Oversampling Extraction. *KIIE, 45(6),* 539-546.

[14] K. J. Jang. (2017). A study on Innovative Financial Services of Business Models using Blockchain Technology. *e-Business, 18(6),* 113-130.

[15] J. H. Jin & K. J. Ko. (2018). Blockchain technology trends and direction of utilization of health and welfare information statistics. *Health and Welfare Forum, 2018(4),* 96-106.

[16] G. S. Jeong. (2017). *Special Report: blockchain-based insurance service trends.* Seongnam : TTA.

[17] A. R. Bologa et al. (2013). Big data and specific analysis methods for insurance fraud detection. *Database Systems Journal, 4(4),* 30-39.

[18] K. Nian et al. (2016). Auto insurance fraud detection using unsupervised spectral ranking for anomaly. *The Journal of Finance and Data Science, 2(1),* 58-75.
DOI : 10.1016/j.jfds.2016.03.001

[19] J. Schiller. (2006). The impact of insurance fraud detection systems. *Journal of Risk and Insurance 73(3),* 421-438.
DOI : 10.1111/j.1539-6975.2006.00182.x

[20] J. S. Choi & J. K. Park. (2017). A Study on Blockchain-based Decentralized Internet of Things Paltform. *KIISC, 27(6),* 5-14.

[21] D. Y. Lee et al. (2017). Key Blockchain Technologies and Trends at Home and abroad. *Communications of the Korean Institute of Information Scientists and Engineers, 35(6),* 22-28.

[22] Y. J. Shin. (2021). The Improvement Plan for Personal Information Protection for Artificial Intelligence (AI) Service in South Korea. *Journal of Convergence for Information Technology, 11(3),* 20-33.
DOI : 10.22156/CS4SMB.2021.11.03.020

[23] Y. S. Jeong et al. (2019). A Blockchain-based IIoT Information Collection Model for Improving the Productivity of Small and Medium Businesses. *Journal of Convergence for Information Technology, 9(12),* 1-7.
DOI : 10.22156/CS4SMB.2019.9.12.001

[24] X. Zheng, Y. Le, A. P. Chan, Y. Hu & Y. Li. (2016). Review of the application of social network analysis (SNA) in construction project management research. *International journal of project management, 34(7),* 1214-1225.
DOI : 10.1016/j.ijproman.2016.06.005

[25] M. K. Nasution. (2016). Social network mining (SNM): A definition of relation between the resources and SNA. *International Journal on Advanced Science, Engineering and Information Technology, 6(6),* 975-981.

[26] J. Jianqiu et al. (2013). Min-max hash for jaccard similarity. *In 2013 IEEE 13th International Conference on Data Mining* (pp. 301-309). IEEE.
DOI : 10.1109/ICDM.2013.119

**이 용 주(YongJoo Lee)** **[정회원]**

- 1999년 2월 : 청주대학교 정보통신공학과(공학사)
- 2001년 2월 : 충북대학교 전자계산학과(이학석사)
- 2019년 2월 : 충북대학교 전자계산학과(이학박사)
- 2001년 1월~2009년 12월: 한국전자통신연구원 선임연구원
- 2020년 9월 ~ 현재: 충북대학교 박사후연구원
- 관심분야 : 블록체인, 정보보안, 기계학습, IoT
- E-Mail : yjlee3363@naver.com