Research Article

A. Oommen Philip* and RA K Saravanaguru

Secure Incident & Evidence Management Framework (SIEMF) for Internet of Vehicles using Deep Learning and Blockchain

https://doi.org/10.1515/comp-2019-0022
Received Sep 16, 2019; accepted Mar 05, 2020

Abstract: Even though there is continuous improvement in road and vehicle safety, road traffic incidents have been increasing over last few decades. There is a need to reduce traffic incidents like accidents through predictive analysis and timely warnings while at the same time data related to accidents and traffic violations need to be maintained in a tamper proof storage system that can be retrieved for forensic analysis and law enforcement at a later stage. The Secure Incident and Evidence Management Framework (SIEMF) proposed in this work address these two challenges of predictive modeling for timely warning and secure evidence management for forensics analysis in case of accidents and traffic violations. The system proposes a deep learning based predictive incident modeling with blockchain and CP-ABE based access control for the incident data stored in blockchain.

Keywords: Internet of Vehicles, Deep Learning, Blockchain, Accident Forensics, Evidence Management

1 Introduction

According to the World Health organization 1.35 million people die each year as a result of road traffic crashes [1]. Traveling over the speed limit (speeding) and non-conformance with traffic rules are the primary reasons for the same. For most accidents, the reporting takes more than an hour and in most cases evidence for liability is not available and cases have to be closed for a lack of evidence. When accidents occur on remote highways where there are no humans to evidence the incident, post collision forensics becomes difficult. Without this information, settlement of cases and insurance also becomes a challenge.

Internet of Vehicles (IOV) is evolving as a new domain of research from VANETS (Vehicular Ad hoc Networks) with the aim of connecting vehicles to the internet [2]. IOV will be a potential solution for accident warning, evidence collection and enable the vision of connected and intelligent transportation in future [3]. IOV allows five types of communication namely Vehicle-to-Vehicle (V2V), Vehicle-to Roadside unit (V2R), Vehicle-to-Infrastructure (V2I), Vehicle-to-Personal devices (V2P) and Vehicle-to-Sensors (V2S), in general termed as V2X communications. V2X communication would enable all vehicles to be connected and be capable of providing vehicular data prior to an accident for accident forensics. Vehicle to road side communication is made possible by RSU (Road Side Units). RSU can be connected to server infrastructure using the internet. With effective coordination between the RSU and the vehicles, the problem of incident warning and evidence management can be addressed.

The SIEMF framework proposed in this work uses a deep learning model which utilises both road and climate conditions as well as vehicle driving patterns as parameters to build an accident warning system for vehicles. The best parameters and driving conditions for a particular road segment or highway at a particular time can be predicted using the deep learning model and subsequently transmitted as a warning message to the vehicles via the RSU. In addition to the learned parameters, traffic rules like speed limit, traffic signal timings and one-way entry restrictions can also be transmitted to vehicles. The vehicles are expected to follow the rules and warnings issued for ensuring traffic safety and any violations made need to be recorded. A Blockchain is a distributed digital ledger that records transactions across many computing nodes so that any involved record cannot be altered without alteration of all subsequent blocks. This allows participants to verify and audit transactions independently. In the framework proposed in this work, data prior to an acci-

***Corresponding Author: A. Oommen Philip:** School of Computing Science and Engineering, VIT, Vellore, Tamil Nadu, India; Email:abinphilip1987@gmail.com
RA K Saravanaguru: School of Computing Science and Engineering, VIT, Vellore, Tamil Nadu, India; Email: saravanank@vit.ac.in

dent and traffic violations can be recorded in a blockchain and evidence can be managed as transactions with all concerned parties creating a record in Blockchain. In order to decide who can add and access the transactions in a Blockchain framework, access based authentication control is proposed over the Blockchain architecture. Since the transactions are stored in public storage, additional cryptographic support is enforced using CP-ABE. (Cipher Policy- Attribute Based Encryption). In case of a dispute, various parties such as law enforcement officials, traffic departments, insurance agencies and vehicle owners can access the framework for post incident analysis.

The contributions in this work are listed below

1. To prevent continuous data transfer from vehicles to RSU, trained traffic daemon is moved as an agent from the RSU to the vehicles. The daemon filters and sends only important information to the RSU. Due to this process both network and storage overheads are reduced.
2. Fine grained access control on vehicular evidence information is ensured using CP-ABE over Blockchain in the proposed solution.
3. Efficient Indexing of vehicular information is maintained by SIEMF Framework, due to which retrieval of data becomes easy and efficient in the proposed solution.
4. Blockchain has been used for managing the keys, evidence information etc. There by integrity of key and evidence is ensured in the proposed SIEMF Framework.
5. Machine learning has been used for predicting incidents and issuing warning messages for vehicles. Blockchain is used for evidence management in case of accident and warning violation. Thus, the paper proposes a framework demonstrating the convergence of Machine Learning, Blockchain and IoT.

The paper is organized as follows. Section 2 details the related works in areas of accident prediction, block chain storage and access. Section 3 details the open issues identified from the survey. Section 4 details the proposed solution in detail. Section 5 presents the results and the comparative analysis of the results. Section 6 presents the conclusion and details of future scope of the work.

2 Related Work

We aim to build a framework by combining Incident prediction using Deep learning and CP-ABE based Blockchain verification by various Stakeholders in case of traffic incident, accidents and traffic violations.

As the framework proposes incident prediction and Blockchain based evidence management, a survey was conducted to study the developments in the two separate categories:

1. Accident Prediction using machine learning;
2. Blockchain-based storage and access.

2.1 Survey on Accident Prediction using Machine Learning

In [4] the authors analyzed the change of traffic flow under fog conditions. A crash risk indicator model was designed based on changes in traffic flow during fog conditions. A binary logistic regression model was applied to link the crash risk to traffic flow characteristics and based on it crash risk was predicted. Crash risk indicator was modeled in terms of traffic flow characteristics like speed, occupancy and traffic flow relationship. The approach did not give importance to the past incidents, which would have been an important parameter for modeling crash risk. In [5] authors made a study to identify and quantify the impact of roadway and environmental factors to traffic crash severities. They analyzed the impact of following parameters like crash location, road alignment, light conditions, speed limit etc. using Logistic regression. They analyzed the effect of each parameter individually on crash risk without analyzing the impact of correlation of the parameters. In [6] authors used different predictive analytics algorithms to model the complex relationship between the crash related risk factors and severity of the injury. The relative importance of crash relative factors is identified using a systematic series of information fusion-based sensitivity analysis. The sensitive analysis was performed at removing each risk factor variable and analyzing the impact on accuracy of incident severity. Relationship between the fatal rate and the parameters of road, environment and driving is analyzed using association rule mining and the variables identified are used to train a naïve Bayes algorithm for classification in [7]. A hybrid approach combining both a multinomial logit model and a Bayesian network method for analysis of driver injury severities in rear end crashes is proposed in [8]. The Multinomial logit model identifies the contributing factors for rear end severities and the Bayesian network formulate statistical association between the injury severity outcome and the contributing factors. Similar to [6], this work does not consider the correlation between the contributing factor for modeling the injury severity. Authors in [9] proposed an incremental data-based crash risk prediction model. A support Vector Machine based crash risk prediction model was trained and then further improved with incremental data. Incremental learning proved to be more efficient than one-time learning. In [10], the author applied the Bimal Ghimire model using big data analytics to predict accidents. Spatial decision tree for the prediction of road accidents is proposed in [11]. Spatial decision tree is a special form of Conventional Decision Tree (CDT) which integrates the spatial dimension within the conventional decision tree. It replaces the measure of entropy in the CDT with the measure of spatial entropy. The number of classes obtained in SDT is less when compared to CDT. This indicates that SDT facilitates the discovery of the most dominant classification rules. Further, for large spatial dataset, SDT is found to be faster than CDT. Authors in [12] proposed an improved deep learning model to explore the complex interactions among roadways, traffic, environmental elements, and traffic crashes. The proposed model includes two modules, an unsupervised feature learning module to identify functional network between the explanatory variables and the feature representations and a supervised fine-tuning module to perform traffic crash prediction. The model could not predict the spatial temporal dynamic pattern in crash data. In [13] authors conducted a thorough analysis on SVM models' predictive capability, the importance of variable selection before developing SVM models; and the effect of the explanatory variables in the SVM models for the case of accident prediction. They concluded SVM model with Radial-basis kernel function outperformed other classifier in predictive ability.

2.2 Survey on Blockchain-based storage and access

Blockchain technology has been proposed in literature as a solution to record transactions that can be verified later by various stakeholders. Authors in [14] secured the cloud storage services using a minimally trusted Blockchain system. It prevents the cloud storage against forking attacks. All file operations are logged to Blockchain to avoid any forking claims. A decentralized storage system using Ethereum Blockchain and Attribute Based Encryption (ABE) is proposed in [15]. The solution was proposed to solve the problem of an un-trusted cloud service provider in the case of Attribute based access control. The cloud service provider can cause key abuse and privacy leakage issue. The keyword search on cipher text is realized using smart contract on the Ethereum Blockchain. A Blockchain based access control framework for big data was proposed in [16]. Authors in [17] proposed a novel Blockchain-based threshold IoT service system called Beekeeper. The communication between server and devices for IOT computations is done as Blockchain transactions. Since all computations are only through Blockchain transactions, the computation history is maintained and keeps the records of any false computations by servers. Since the computation records in Blockchain is distributed, availability also increases. Blockchains are computationally expensive and involve high bandwidth overhead and delays. This makes it, unsuitable for energy deficient IOT devices. A light weight Blockchain-based architecture eliminating the overhead of classic Blockchain without compromising the security and privacy of Blockchain was proposed in [18]. The ledgers are managed centrally, and distributed trust is adopted to reduce the block validation processing time. Access control for IOT based on Blockchain with decentralized pseudonymous was proposed in [19]. The access control between the requestor and the resource (IOT device) is encoded as a token and stored in Blockchain. When access is made by the requestor, validation is done by referencing the Blockchain and access control enforced. In [20] authors used Blockchain for ensuring data integrity for data stored in the cloud. Third Party Auditors (TPA) are usually adopted for enforcing integrity of data stored in semi trusted clouds, but if the TPA is compromised, integrity is at risk. In [20], the hash code for the file stored in cloud is maintained in Blockchain and validation is done at owner end without relying on Auditor. Block chain based secure distributed data storage with keyword search service is proposed in [21]. Mobile applications installed on a user's phone can use personal data for service personalization. [22] proposed a way to manage this access by a third party, by offloading user data to Blockchain and maintaining a transaction every time service accesses the data. In this way a user can own the data and control its access. In [23], the author proposed a permissioned Blockchain framework for managing the collection of vehicular data for forensic analysis of traffic accidents. They used vehicular public key infrastructure (VPKI) in a permitted blockchain and a fragmented ledger, which enabled storage of hashed data in the shared ledger while the details could be stored in fragmented ledgers as non hashed data. In addition, the use of pseudonyms for identities helped preserve the privacy of users.

3 Literature Gap

From the survey we identify the following shortcomings in existing solutions:

1. Most systems depend on passing of vehicular information to RSU frequently without paying attention to the importance of the information. Due to this the network and storage overhead for storing this information on the server side increases.
2. Most machine learning models considered instantaneous data alone for incident classification. Not much relevance has been given to continuous data over past intervals.
3. Fine grained access control and secure sharing of vehicular information related to forensics among various stakeholders has not been considered.
4. Efficient retrieval of vehicular forensic information from Blockchain by various stakeholders has not been considered.
5. No work discusses combining Machine Learning and Blockchain to address the issue of Incident Prediction and Evidence Management in the Internet of Vehicles.

4 Proposed SIEMF Framework

Usage of Vehicle generated data for Vehicular forensics has been a topic of interest for researchers [24]. It helps forensic investigators in finding supporting evidence from the digital data generated by vehicle and its nearby vehicles at the time of an incident. One important objective is that evidence collected during the incident must be kept in an immutable secure storage that can be referred to at any time by stakeholders. The insurance company may need to analyze the evidence for liability assessment and decision making. Law enforcement agencies may need to refer to it at later point for making decisions based on law and consequently delivering a verdict. The Roadway department

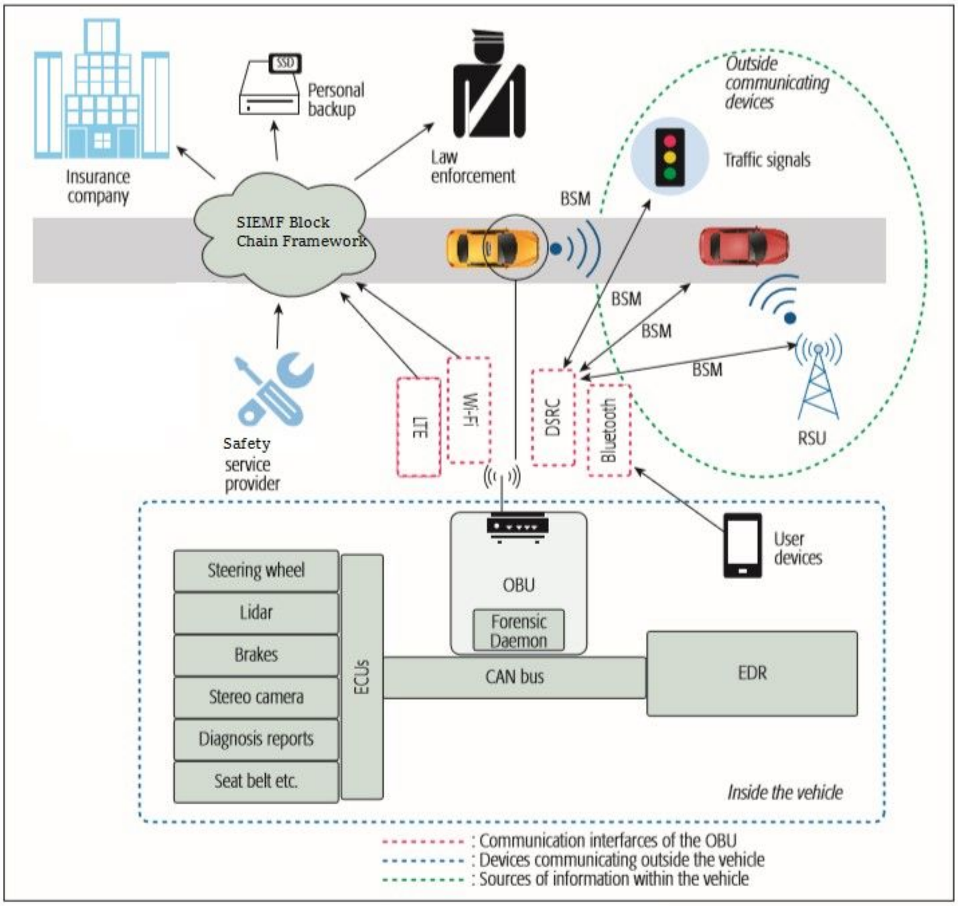

Figure 1: Interaction between SIEMF framework, Vehicles and stakeholders (used from [23])

may refer it to improve road conditions for safe driving and identify accident prone zones. Vehicle servicing companies may need to access the details to analyze a vehicle fault. With the availability of wireless communication infrastructure (DSRC, Wi-Fi, Bluetooth, LTE) safety warning and traffic regulations like speed, traffic light timings, restricted entry etc. can be issued to vehicles to ensure traffic safety and prevent accidents. Vehicles can be tracked and this information can be used to personalize the safety warning. SIEMF platform has been designed with above requirements without compromising on the security and privacy of vehicles with deep learning and blockchain technology as the enablers. SIEMF framework is build on top of existing VPKI infrastructure. Building the SIEMF framework on top of VPKI has following advantages:

1. Anonymity of communicating parties;
2. Resiliency against privacy and spoofing attacks.

The architecture showing the interaction between SIEMF framework, vehicles and Stakeholders is shown in Figure 1. The interaction framework architecture is inspired from [23]. Vehicles communicate via RSU. Each vehicle is equipped with an On-Board Unit (OBU) which has Forensic daemon running on it. The Forensic daemon reads the data continuously from the CAN bus, which is the backbone of the vehicle network. The CAN bus deliver valuable data in terms of vehicular forensics to the OBU that can be retrieved by the forensic daemon. The road conditions and weather data are continuously delivered by RSU to the vehicles. This data is also collected by the forensic daemon. Safety Service Provider (SSP) issues an accident warning and speed alerts to drivers via RSU. Safety Service Provider uses deep learning model to provide safety warning based on the Road conditions, Environmental conditions and Driving conditions. We train a Convolutional Neural Network (CNN) and the trained CNN model object is delivered as serialization object to the OBU to deliver continuous safety warning to the vehicles. Thus, the classification and warning generation occurs at the vehicle side incorporating the vehicle and environment parameters. At the time of any incident or traffic violation, transaction data about incident or violation is created and sent to RSU for updating to the SIEMF Blockchain framework.

Three salient features of SIEMF framework are:

1. Deep Learning based Safety Service Provider;
2. CP-ABE based evidence access control;
3. Blockchain based evidence management.

4.1 Deep Learning based Safety Service Provider

Most of the incident warning systems in literature create a prediction model between the incidents and contributing factors like road conditions, environment conditions and driving conditions. The model is then used to predict the incidents and warn the vehicles. The models are based only on instantaneous data and not on continuous data. Usage of instantaneous data for modeling has following drawbacks

1. Due to rapid change in parameters, there would be frequent conflicting warnings
2. The cause of incident would have occurred sometime back before the incident is reported. This factor cannot be accommodated in the model.
3. Model lacks adaptive and continuous learning ability

Continuous data based predictive modeling with Adaptive Deep Learning is proposed as the solution for the above three problems in this work. The authors intension is not to propose an efficient deep learning model, but to show the feasibility of training a model based on continuous data specific to a region. The architecture of the Deep Learning Convolutional Neural Network (CNN) model is given in Figure 2. The data collected over N consequent intervals is assembled as matrix and used as input for a Convolutional Neural Network.

A Loop controlled Convolutional Neural Network is trained to obtain the label for the input matrix M. The input parameters used to train the model is listed in appendix Table 1. The CNN model is a three-layered feed forward network using radial bias function. At hidden layer, Gaussian Activation function is applied and at output layer linear activation function is applied in the proposed CNN model.

The prediction output of the neural network classifier with K hidden layer neuron is given as

$$y_j^i = \propto_{j0} + \sum_{k=1}^{K} \propto_{jk} \Phi_k \left(X^i\right), \quad j = 1, 2 \ldots n$$

The parameters in above equation is detailed in Table 1.

The parameters for fine tuning the performance of the convolutional neural network is modified using Loop control mechanism. Loop controlled fine tuning is based on two measurements:

1. Estimated Class label (P1)
2. Maximum Hinge Error (P2).

Estimated Class label is the label of the maximum firing output layer neuron. The Hinge Error is calculated as

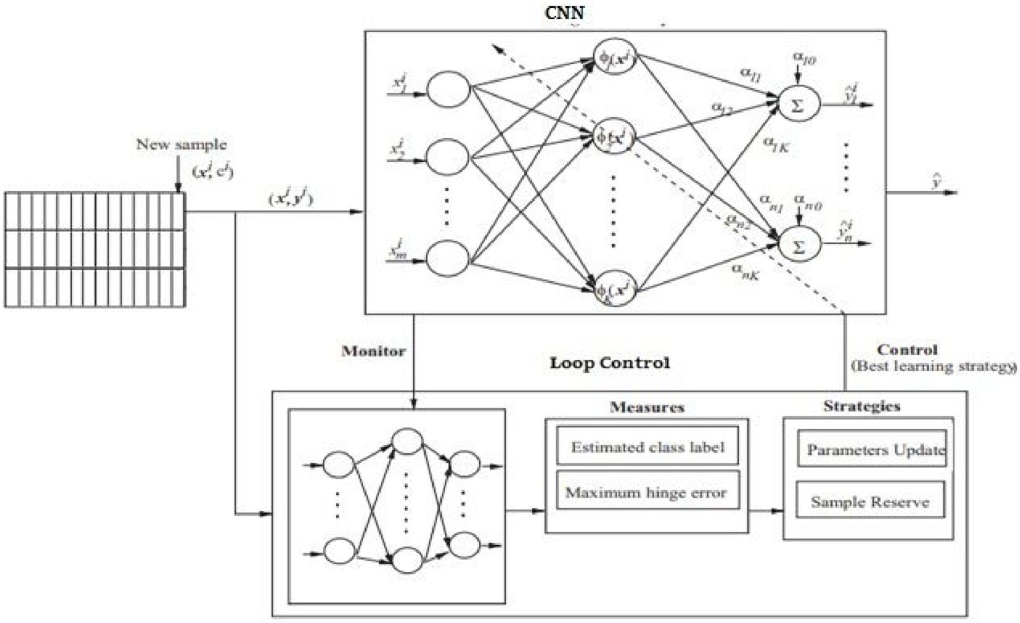

Figure 2: CNN Model

Table 1: Hidden layer Neural parameters

N	number of output layer neurons
$\propto_{j0}$	basis to the *j* output neuron
$\propto_{jk}$	Weight connecting *j* to *k* th neuron
$\Phi_k\left(X^i\right)$	The response of k th hidden neuron to input modeled as $\Phi_k\left(X^i\right) = \exp\left(-\frac{\lvert\lvert x^i - \mu_k \rvert\rvert^2}{\sigma_k^2}\right)$ represent the center and width of kth hidden neuron

the error between predicted and the actual result. The maximum of the error across all data sample is the Maximum Hinge Error. The training data is split to different samples and the parameters (P1, P2) are measured for subset of dataset. The parameters of hidden layer neuron and the bias are modified until satisfactory values for parameters P1, P2 are obtained.

The Deep learning classifier is trained to predict the output label (incident type as 0 or 1) based on three contributing factors of Road conditions (road type, location, alignment, surface, speed limit, traffic density) Environmental conditions (weather, light, rain) and Vehicle conditions (vehicle type, engine, break status). The model can also be modified to contain multiple classes.

The trained CNN model object is moved to each vehicle via RSU and it continuously classifies the incident type to 0 (Normal) or 1 (Warning).

RSU trains the CNN model specific to the conditions of the area which it monitors. Say for a school zone area, the vehicle speed above 30 km/hr may be considered as a violation, the CNN model will be trained to detect speed above 30 km/hr as a violation. The classification of incident type takes place in the vehicle continuously based on the contributing factors.

4.2 CP-ABE based Evidence access control

Accidents, rash driving, collision scenarios, traffic violations are considered as incidents. When an incident occurs, the Forensic daemon collects all information from CAN Bus and the incident warning reports issued in last N duration and sends this information to RSU. Vehicle to RSU communication is authenticated using VPKI, thus it prevents forging of information. In this way, only valid messages can reach RSU. The information transmitted by Forensic daemon to RSU is complete in the sense that it includes lot of information generated and stored by the OBU. The access to all the information need not be provided to all the stakeholders. RSU decides the access level of each attribute. Fine grained access control is enforced for stakeholder using Blockchain modified with CP-ABE.

The Blockchain is a distributed database that records all the transactions that have occurred in the peer-to-peer network. All participants in the network hold the same copy of the ledger.

In this network, there is no central authority and there is no single node that can control the entire network. Blocks are added to the blockchain by a consensus among most nodes in the system. Each block contains a block header and a series of transactions, each block header contains the link pointers of the block headers of the previous block, the Merkle root of the tree-like transaction information, and a timestamp. In this way, the blocks are linked together in chronological order. The cryptography hash algorithm ensures that the transaction data in each block is immutable, and the linked blocks in the blockchain cannot be tampered.

The original CP-ABE algorithm [25] is modified to work on top of Blockchain platform. The CP-ABE has following 4 steps

Setup – This stage generates the public parameters (PK) and a master key (MK)

Encrypt (PK, M, A) – The encryption algorithm takes as input PK, a message M, and an access structure A over the universe of attributes. The algorithm will encrypt M and produce a ciphertext (CT) such that only a user that possesses a set of attributes that satisfies the access structure will be able to decrypt the message. We will assume that the ciphertext implicitly contains A

Key Generation (MK, S) – The key generation algorithm takes as input the master key MK and a set of attributes S that describe the key. It outputs a private key SK

Decrypt (PK, CT, SK) – The decryption algorithm takes as input the public key PK, a ciphertext CT, which contains an access policy A, and a private key SK, which is a private key for a set S of attributes. If the set S of attributes satisfies the access structure A then the algorithm will decrypt the ciphertext and return a message M.

The SIEMF Framework flow is illustrated in Figure 3. The work flow of the SIEMF framework is explained below and continued in section 4.3. The SIEMF runs the setup one time during initiation and creates two Blocks in Blockchain. One block is between the SIEMF framework and the RSU. This block has information about encrypted PK and access attribute structure for each stakeholder. The encryption is done using symmetric cryptographic algorithm with key shared between the SIEMF framework and the RSU. Another block is between SIEMF framework and the stakeholders. It has the information of encrypted secret key SK for the stakeholder generated using Key Generation function.

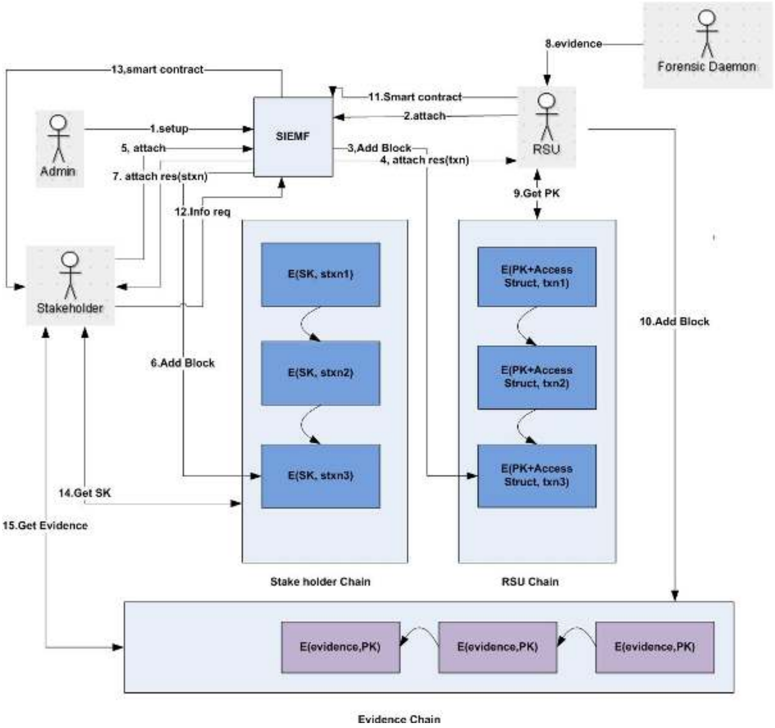

Figure 3: Proposed SIEMF Framework flow

The encryption is done using symmetric cryptographic algorithm with key shared between the SIEMF framework and each of the stakeholder. Each of the blocks are in two separate Block chain say RSU-Chain and Stakeholder-chain. Every time a new stakeholder is added or removed, or attribute access is modified, new block is added in the Stakeholder chain. Similarly, for each new RSU, a block is added in RSU Chain.

4.3 Block Chain based Evidence Management

Every time RSU receives evidence information from Vehicle in its coverage area, it encrypts the evidence using the PK and Access attributes by referring to the corresponding Block for RSU in the RSU-Chain. The encrypted evidence information is then added to Block chain between RSU and all stakeholders (Evidence Chain). Every time a stakeholder wants to access the evidence information, they retrieve the secret key SK from the block in the Stakeholder chain and uses the SK to decrypt the evidence. The attributes which are given control to that stakeholder can be decrypted and accessed by the stakeholder. The flow of the entire process from setup, creating evidence and retrieval of evidence is shown in SIEMF Flow (Figure 3).

There are three different Blockchain:

1. Stakeholder chain;
2. RSU chain;
3. Evidence chain.

Stakeholder chain maintains the secret key information for decrypting the evidence information stored in the evidence chain. For each stakeholder a secret key is generated and transaction with encrypted secret key is added in the stakeholder chain, the transaction reference id is generated for the transaction and shared to the stakeholder. Through the transaction reference id, the stakeholder can retrieve the block and decrypt the secret key. By keeping the secret key in encrypted form, it is protected against any key inference attacks. Every time a RSU attaches to SIEMF framework, the encrypted public key is added to RSU chain as transaction and the transaction reference id is shared to the RSU. RSU retrieves the public key using the transaction reference and use this key for encrypting the evidence information.

Forensics daemon running within the OBU continuously probes the data collected from the sensors and passes the data to the CNN model. When CNN model detects an abnormality, the evidence data is sent to RSU. RSU encrypts the evidence using the public key retrieved from the RSU Blockchain, and transaction is added to evidence block chain. A smart contract with reference to the transaction is created and stored in the SIEMF framework. Stakeholder can retrieve the related transaction in the evidence chain using keyword-based retrieval.

SIEMF framework maintains the index and mapping keywords in addition to the transaction reference id in the evidence block chain. Stakeholders can query any evidence using keyword and initiate the search. Examples for keyword-based search include searching using vehicle number, road segment etc. The SIEMF framework looks up the keyword and returns the matching transactions in form of a smart contract. The smart contract has the encrypted transactions matching to the keyword searched by the stakeholder. Once the stakeholder receives the smart contract, each encrypted transaction reference id is decrypted. Stakeholder retrieves the transaction from Blockchain and decrypts using secret key.

Smart contract is the agreement between the stakeholder and the SIEMF framework for passing the matching transaction corresponding to the stakeholder's query. The smart contact has all information in encrypted form, so only a valid stakeholder can decrypt and retrieve the transactions from the evidence block chain.

By ensuring the smart contract information is in encrypted form, the probability of launching denial of service attack with the Smart contract on the evidence block chain is averted.

Even if attackers get access to the transaction in the stakeholder chain, RSU chain or Evidence chain, the information is encrypted and attacker cannot make use of it and thus security of the transaction is ensured.

5 Results

The SIEMF framework was tested for a sample VANET network created using a SUMO simulator. For the VANET simulation, a 1000m one-way road segment with four lanes was created in the traffic simulator SUMO. 100 vehicles including buses, sedan cars, and vans enter the road highway segment from both ends of road per minute and leave the simulation after reaching the end of the road. 4 different speed limits are set on each of the lane. RSU was deployed in each lane for every 250m.

The dataset used for the study is synthesized. The purpose of generating a dataset was to demonstrate the feasibility of event prediction and violation scenarios. As mentioned, the authors intension is to show the feasibility of training a model based on continuous data specific to a region and subsequent evidence management using block chain in case of an incident. Thus, we train a model specific to a region. The parameters used as input for training is mentioned in Table A1. Each RSU can have a deep learning model trained specifically for the spatio temporal requirements in that region. Once an incident is detected in the form of ignoring the warning messages or an accident like scenario, it is written as a transaction onto the Blockchain, and can be verified by various stakeholders. The generated dataset was split into 80:20 ratio and 80% was used for training the CNN model. Each vehicle in the simulation used the CNN model to predict the incident. The performance of the CNN is measured in terms of standard parameters: Sensitivity, Specificity, False Positive rate, Precision, F-Measure and Accuracy.

The proposed CNN model is trained with specifications as given in Table 2.

Table 2: CNN Specifications

Input Layer Neurons	19*5 [values of 19 features sampled at every 5 seconds for 5 intervals]
Hidden Layer Neurons	191
Output Layer Neurons	2
Input and Hidden layer activation	Relu
Output layer activation	Sigmoid
Optimizer	Rmsprop

The flow followed for training the proposed CNN model is given below in Figure 4

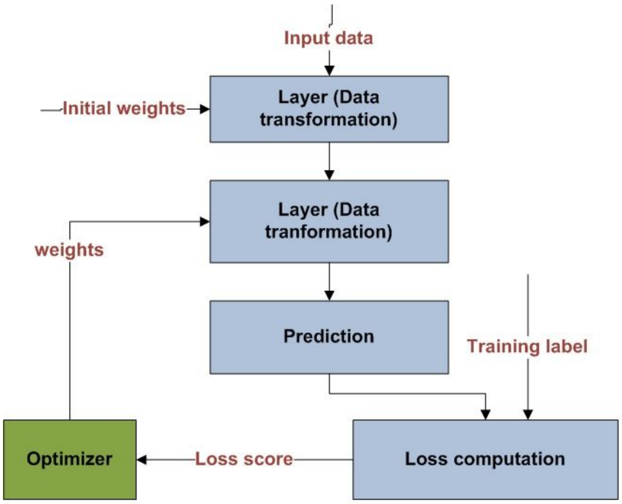

Figure 4: CNN Optimization flow

The results are compared with [8, 9] in terms of sensitivity, specificity, false positives, precision and F-Measure and shown. Although their work was proposed on a different dataset. The model is evaluated on the data set generated using the input parameters mentioned in Table A1 for comparing with our proposed CNN model.

The sensitivity is higher in the proposed model due to the use of multiple features in constructing the classification model (Figure 5).

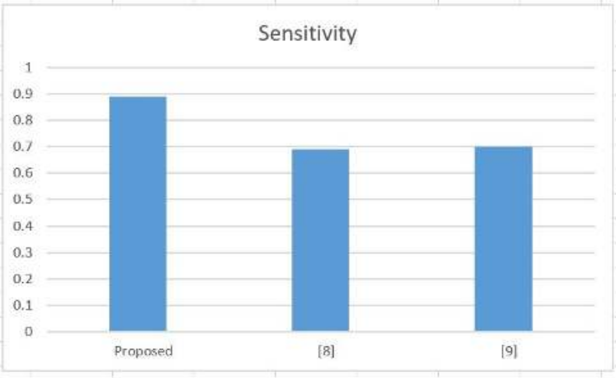

Figure 5: Sensitivity comparison

The specificity is also higher in the proposed solution due to use of adopted loss function for error correction (Figure 6).

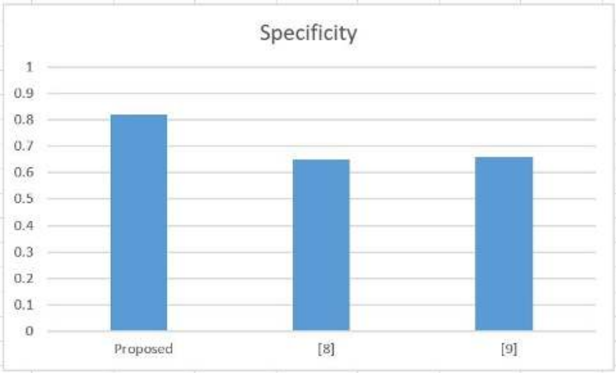

Figure 6: Specificity comparison

Due to the continuous learning and incremental learning used in the proposed classification, the false positive rate is reduced by a large scale if compared to [8] and [9], (Figure 7).

Due to use of best choice of hinge loss function in learning, the precision has increased in the proposed solution compared to [8] and [9], (Figure 8).

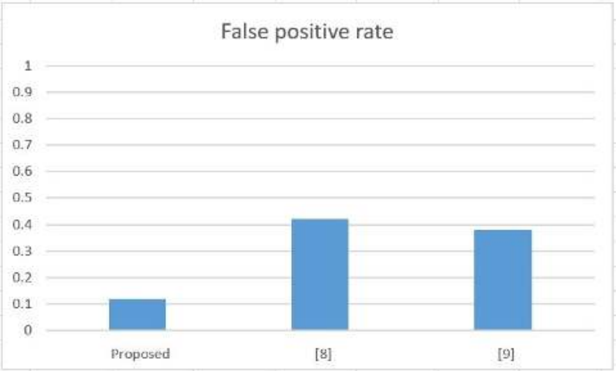

Figure 7: False positive rate comparison

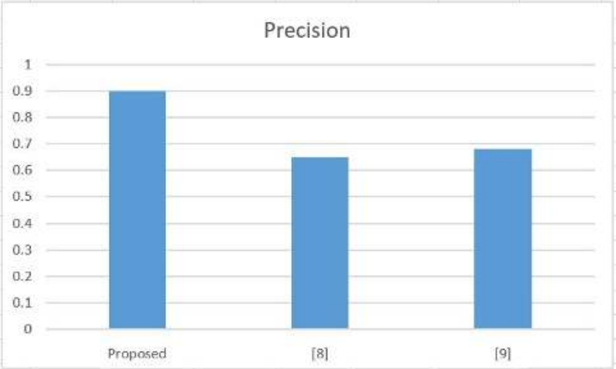

Figure 8: Precision comparison

The balance of precision and recall measured in terms of F-Measure is also higher in the proposed solution when compared to [8] and [9] proving the effectiveness of classification process in the proposed model, (Figure 9).

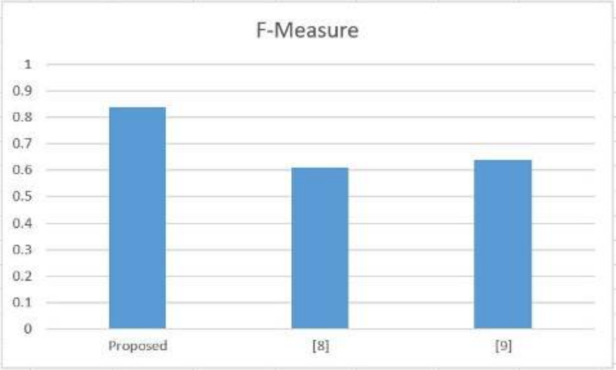

Figure 9: F-Measure comparison

From the results, it can be seen that the proposed CNN model is able to perform better due to continuous training model instead of instantaneous data alone.

For the same set of features, the accuracy is measured for different supervised machine learning classifiers and the result is shown in Figure 10.

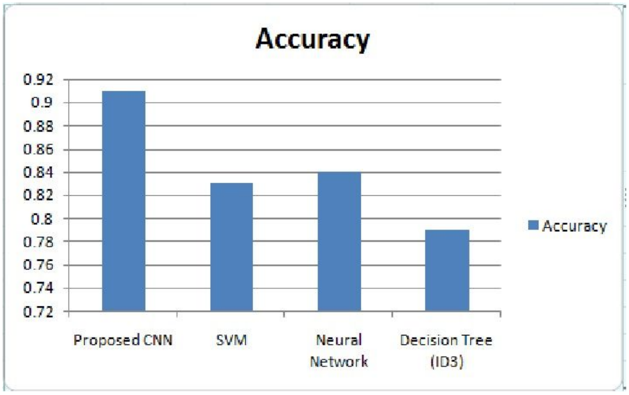

Figure 10: Classifier accuracy comparison

The accuracy in the proposed CNN is due to the loop control strategy adopted in the proposed CNN model. The accuracy of proposed CNN is measured for three different hinge functions in loop control. The hinge loss functions used for experimentation is given in Table 3.

Table 3: Hinge loss functions

Hinge Loss function	Objective function
Crammer and Singer	$\max\left(0, 1 + \underset{y \neq t}{\max} w_y x - w_t x\right)$ where t is the class label w_t and w_y are the model parameters
Weston and Watkins	$\sum_{y \neq t} \max\left(0, 1 + \underset{y \neq t}{\max} w_y x - w_t x\right)$
Zhang quadratically smoothed	$\begin{cases} \frac{1}{2\gamma}\max(0, 1 - ty)^2 \forall ty \geq 1 - \gamma \\ 1 - \frac{\gamma}{2} - ty, \quad otherwise \end{cases}$

The accuracy of proposed CNN is measured for above three hinge loss function and the result is given below.

The summary of proposed CNN results is given in Table 4.

Proposed, CNN has achieved good validation accuracy with high consistency, (Figure 12, 13).

Ethereum Blockchain is used in this work. The parameters for CP-ABE keys is given in Table 5.

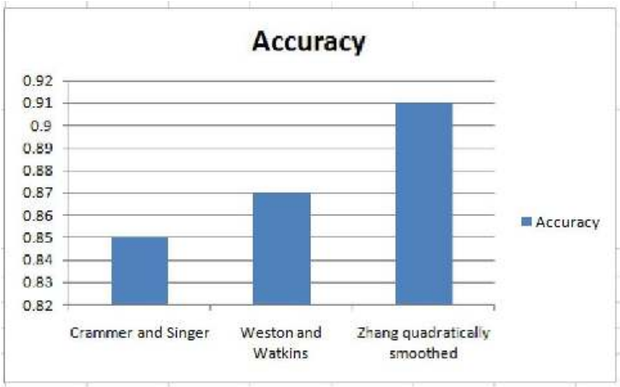

Figure 11: Hinge loss accuracy comparison

Table 4: CNN results

Time/Epoch	48 sec
Training accuracy	96
Validation accuracy	91

Table 5: CP_ABE Key Size

System master Key Size	208 bytes
User secret key	1088 bytes
Ciphertext	Varied from 1 to 100 kb

The performance of Blockchain based implementation is analyzed in terms of block write time, block read time and ability to retrieve data as a number of evidence generated blocks increases over time. The performance is evaluated against the block write, read and retrieval time of [15]. In our model a block consists of a single transaction generated by the RSU corresponding to an incident. This is in contrast to a block having multiple transactions in Block chain. Thus, the transaction write and read time is less in comparison to the compared model.

Block write time in this proposed solution is lower than that of [15]. The reason for reduced Block write time is due to reduction in the volume of information per block and the cost of encryption. The fluctuation in write time is due to the load on Ethereum Blockchain, (Figure 14).

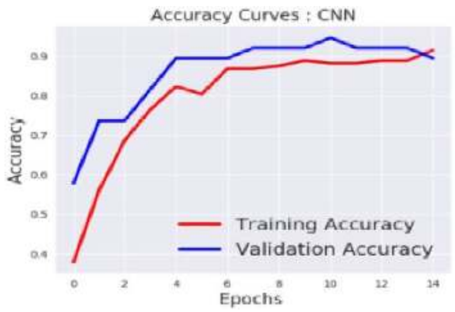

Figure 12: CNN Accuracy

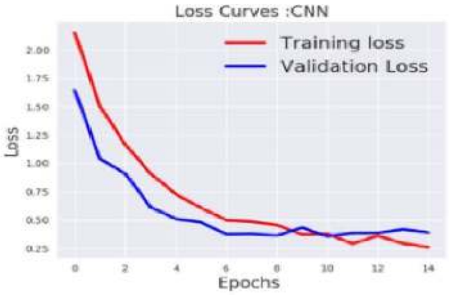

Figure 13: CNN Loss

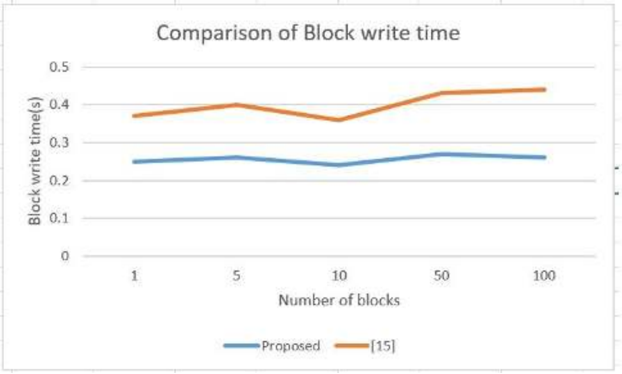

Figure 14: Block write time comparison

Block read time is lower in this proposed solution when compared to [15]. The reason for reduction in block read time is due to the indexing method adopted for the look up of relevant information, (Figure 15).

The time to retrieve the corresponding transaction in the evidence block chain for different rates of abnormal evidence generation rate is measured and plotted in Figure 16.

The retrieval time is bound to an average of 127 milliseconds and is not varying much due to usage of keyword-based hashing followed for indexing the evidence chain transactions in the SIEMF Framework. But in [15] as the transaction increases the retrieval time also increases exponentially.

The time (milliseconds) for storing and fetching blocks for the three different block chains in the proposed SIEMF framework for case of Ethereum block chain is plotted in Figure 17.

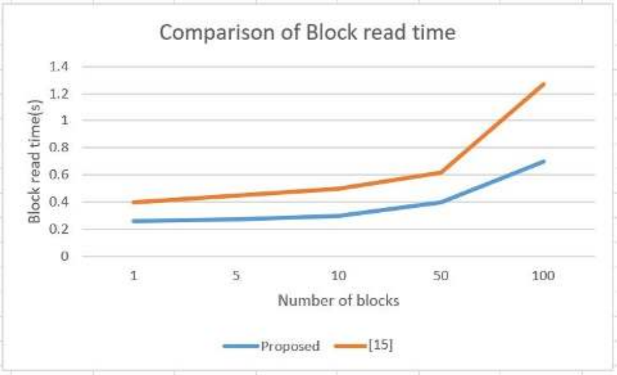

Figure 15: Block read time comparison

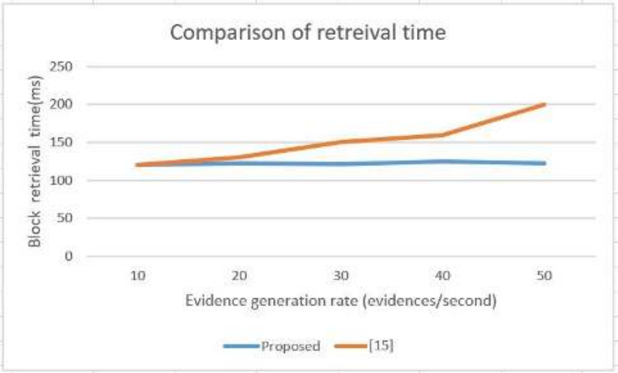

Figure 16: Block retrieval time comparison

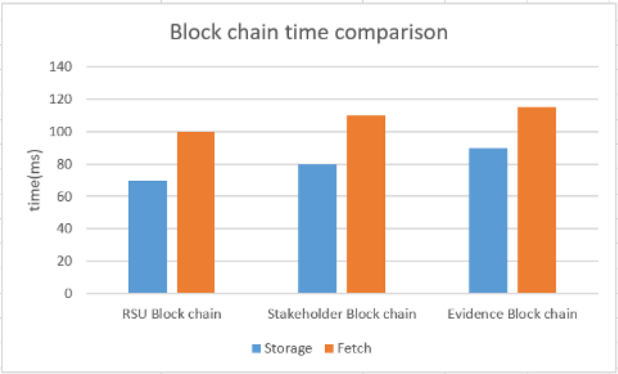

Figure 17: Block store fetch time comparison

6 Discussion

A SIEMF Framework for secure incident and evidence management is proposed in this work. We have highlighted the possibility of training RSU's that are specific to both the spatial and temporal environment and driving conditions, demonstrated the feasibility of modeling a forensic daemon that classifies data within the OBU and finally screens the data send to RSU. The evidence collected from vehicles in case of an accident or warning violation is stored securely in Blockchain. RSU validates the evidences collected from vehicles before uploading to Block chain. CP-ABE based access control is enforced on stakeholders before accessing the evidences. Since the proposed SIEMF Framework is implemented over VKPI which uses pseudonyms for vehicles, the privacy of vehicles (and their owners) is maintained. In future as all vehicles become connected and autonomous they would be automatically transmitted different parameters and driving conditions to be followed as it enters a particular driving zone. These parameters can be trained specific to a zone and any incidents or violations made can be recorded and verified using the proposed SIEMF framework. In case of accidents the data prior to the accident along with the warning messages issued can be recorded on the Blockchain as a transaction.

As a future extension, the authors are working in terms of

1. Developing a suitable consensus mechanism for the vehicles and RSUs to add a block to the block chain.
2. Studying the feasibility of implementing the vehicular evidence frame work using IOTA Tangle which is a Directed Acyclic Graph (DAG) based distributed ledger. Tangle is argued to be more scalable and requires no mining fee when compared to Block chain.
3. Developing better Smart Contract codes specific to Internet of Vehicles and Accident Forensic scenarios incorporating Machine Learning.

References

[1] https://www.who.int/newsroom/factsheets/detail/road-traffic -injuries. Accessed on July 1, 2019

[2] Kaiwartya O., *et al.*, "Internet of Vehicles: Motivation, Layered Architecture, Network Model, Challenges, and Future Aspects," in IEEE Access, 2016, vol. 4, pp. 5356-5373

[3] Philip A O., Saravanaguru RA.K., "A Vision of Connected and Intelligent Transportation Systems", International Journal of Civil Engineering and Technology, Volume 9, Issue 2, February 2018, pp. 873–882

[4] Wu Y., Abdel-Aty M., Lee J., "Crash risk analysis during fog conditions using real-time traffic data, Accident Analysis & Prevention", May 30 2017, Volume 114, pp. 4-11

[5] Wang Y., Zhang W., "Analysis of Roadway and Environmental Factors Affecting Traffic Crash Severities", Transportation Research Procedia, 2017, Vol 25, pp. 2124-2130

[6] Delen D., Tomak L., Topuz K., Eryarsoy E., "Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods", Journal of Transport & Health, March 2017, vol 4, pp. 118-131

[7] Li L., Shrestha S., Hu G., “Analysis of road traffic fatal accidents using data mining techniques”, IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), London, 2017, pp. 363-370

[8] Chen C., Zhang G., Tarefder R., Ma J., Wei, H., Guan H., “A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes”, Accident Analysis and Prevention, 2015, vol 80, pp. 76–88

[9] Sun P., Guo G., Yu R., Traffic crash prediction based on incremental learning algorithm, 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, 2017, pp. 182-185

[10] Park S-h., Kim S-m., Ha Y-g., “Highway traffic accident prediction using VDS big data analysis”, The Journal of Supercomputing, July 2016, Volume 72, Issue 7, pp. 2815–2831

[11] Ghimire B., Bhattacharjee S., Ghosh SK., "Analysis of Spatial Autocorrelation for Traffic Accident Data Based on Spatial Decision Tree”, Fourth International Conference on Computing for Geospatial Research and Application, 2013, San Jose, CA, pp. 111-115

[12] Dong C., Shao C., Li J., Xiong Z., “An Improved Deep Learning Model for Traffic Crash Prediction”, Journal of Advanced Transportation. 2018, pp. 1-13

[13] Yu R., Abdel-Aty M., “Utilizing support vector machine in realtime crash risk evaluation”, Accident Analysis & Prevention, 2013 vol. 51, pp. 252–259

[14] Tang et al., "ChainFS: Blockchain-Secured Cloud Storage", 2018, IEEE 11th International Conference on Cloud Computing (CLOUD), San Francisco, CA, 2018, pp. 987-990

[15] Wang S., Zhang Y., Zhang Y., "A Blockchain-Based Framework for Data Sharing with Fine-Grained Access Control in Decentralized Storage Systems," in IEEE Access, 2018, vol. 6, pp. 38437-38450

[16] Es-Samaali H., Outchakoucht A., Leroy J P., “A blockchain-based access control for big data,” Int. J. Computer Networks Communication and Security, July 2017 vol. 5, no. 7, pp. 137–147

[17] Zhou L., Wang L., Sun Y., Lv P., "BeeKeeper: A Blockchain-Based IoT System with Secure Storage and Homomorphic Computation," in IEEE Access, 2018, vol. 6, pp. 43472-43488

[18] Dorri A., Kanhere S. S., Jurdak R., “Towards an Optimized BlockChain for IoT,” IEEE/ACM Second International Conference on Internet-of-Things Design and Implementation (IoTDI), 2017, Pittsburgh, PA, pp. 173-178

[19] Ouaddah A., Elkalam A. A., Ouahman A. A., “Towards a novel privacy-preserving access control model based on blockchain technology in IoT”, In book: Europe and MENA Cooperation Advances in Information and Communication Technologies Springer, Sept 2017, pp. 523–533

[20] Liu B., Liu B., Yu X. L., Chen S., Xu X., Zhu L., "Blockchain Based Data Integrity Service Framework for IoT Data", IEEE International Conference on Web Services (ICWS), 2017, Honolulu, HI, pp. 468-475

[21] Do H. G., Ng W. K., "Blockchain-Based System for Secure Data Storage with Private Keyword Search", IEEE World Congress on Services (SERVICES), 2017, Honolulu, HI, pp. 90-93

[22] Zyskind G., Nathan O., Pentland A S., "Decentralizing Privacy: Using Blockchain to Protect Personal Data", IEEE Security and Privacy Workshops, 2015, San Jose, CA, pp. 180-184

[23] Cebe M., Erdin E., Akkaya K., Aksu H., Uluagac S., "Block4Forensic: An Integrated Lightweight Blockchain Framework for Forensics Applications of Connected Vehicles", in IEEE Communications Magazine, October 2018, vol. 56, no. 10, pp. 50-57

[24] Nilsson D. K., Larson U. E., “Conducting Forensic Investigations of Cyber Attacks on Automobile In-Vehicle Networks”, Proceedings of the 1st international conference on Forensic applications and techniques in telecommunications, information, and multimedia and workshop, Jan 2008, article 8, pp. 1-6

[25] Bethencourt J., Sahai A., Waters B., “Ciphertext-Policy Attribute-Based Encryption”, IEEE Symposium on Security and Privacy (SP ’07), 2007, Berkeley, pp. 321-334

Appendix

Table A1: Parameters used as input for training CNN Model

Category	Name	Values
Road Conditions	Type	0 – urban major arterial
		1 – rural interstate
		2 – rural minor arterial
		3 – rural major arterial
		4 – urban minor arterial
		5 – urban interstate
		6 – urban other freeway
		7 – highway
	Location	0-not at intersection
		1-at intersection
	Alignment	0-straight
		1-curve
	Surface	0-dry
		1-wet
	Speed Limit	40-200 km/hr
	Traffic Density	0-100%
Environmental Conditions	weather	0-clear
		1-cloud
	Light	0-daylight
		1-dark
		2-dark but lighted
	Rain	0-light
		1-medium
		2-heavy
Driving Conditions	speed	0 to 200 km/hr
	Alcohol detector result	0-no
		1-yes
	Seat belt status	0-no
		1-yes
	Day of travel	0-SUN,1-MON 6-SAT
	Time of travel	0 to 24 hr
	Month of travel	1 to 12
Vehicle Conditions	Type	0-sedan
		1-suv
		2-tempo traveler
		3-Heavy vehicle
	Indicator status	0-working
		1-Not working
	Engine status	0-Healthy
		1-Not Healthy
	Breaking status	0-Healthy 1-Not Healthy