Insurance Fraud Detection using Spiking Neural Network along with NormAD Algorithm

Gopikrishna Panda[1], Sunil Kumar Dhal[1], Rabinarayan Satpathy[1], Subhendu Kumar Pani[2]

[1]Sri Sri University, Cuttack Odisha, India
[2]Krupajal Computer Academy, Bhubaneswar, Odisha, India
Corresponding email: -ergopipanda@gmail.com

Abstract- General automobile insurance in recent years, has seen a huge escalation of fraud cases. The requirement of utilizing well organised and coherent technique to check on or determine user those are potential frauds. Thus, the deployment of the NormAD algorithm with less delay to enhance the safety and authorized in the operative process. The paper here describes attribute extrication method and Spiking Neural Network structure to resolve the issue of identification of automobile insurance fraud. The attribute second-level extrication algorithm coined in this paper can efficiently derive key attributes and enhance the identification accuracy of succeeding algorithms. So as to achieve to resolve the issue of unstable simulation allotment in the automobile insurance fraud identification scheme, an exemplary distributed method established on the plan of small unit proportion balance is presented. Formulated on the above techniques of attributes extrication and sample division, a model established on Spiking Neural Network with NormAD Algorithm is proposed. This method utilizes the complete goal of implementation of the Spiking Neural Network model algorithm that rely on Spiking Neuron, and ultimately accomplishes in enhancing the exactness of the detection of Automobile Insurance Fraud.

Keywords: - Attribute Extrication; Spiking Neural Network; Normalized Approximate Descent (NormAD); Algorithm; Automobile insurance Fraud Identification.

1. Introduction

In various domains, we encounter fraud regularly. It is found in many various moulds and models coming from yesteryears fraud or scams e.g., simpleton like tax fraud, to be explicit, in which whole lot of people in group come together to perform such scams. So, these organised groups are readily found in the automobile insurance domain. Scamster creates accidents in traffic and apply for false insurance claims to profit (injudicious) currency from their general or vehicle insurance. Sometimes it is observed that there are no accidents in actual but the vehicles are located on the road to create false claim for insurance money. However, many insurance claims are unplanned but mere opportunity to make increased claim for covering past car expense. Fake accidents have various common features. These accidents happen in near to midnight and areas related to rural where there are no one to witness the accident or staged people can be used. Usually younger males are the divers, as many passengers are present in the vehicles, excluding children or elders. To validate the whole scene police are present to create the substantial credibility for making false claim easily. The common thing in all this is that total people have several wounds (not serious), whereas this is found mostly vehicles are undamaged. Plenty other sceptical features exist, not defined here. The insurance companies are most fond of groups of scamsters that are organised are as such drivers, chiropractors, garage mechanics, lawyers, police officers, insurance workers and others. These categories are related to major leakage in financial loss.

In works in the literature, different techniques are reflected for determining false claims in automobile insurance domain. However, in various domain key factor is the database, a thin line of work or research is done in the domain of fake claim in insurance fraud detection is present in the databases.

This fraudulent behaviour in automobile insurance is determined by using Latent Dirichlet Allocation (LDA) based text analytics as proposed in [1]. In work [2], genetic algorithm based fuzzy c-means clustering has been coined to standardised the automobile insurance scam in which different supervised classifier structures are used for identification. Another system of multiple classifier [3] established on principal component analysis, random forest and mighty nearest neighbourhood method has been put forward to evaluate the fraudulent tasks in the automobile insurance scam which yields good proficiency than the state-of-the-art models. Various feature selection techniques based on correlation and genetic algorithm [4] has been utilized and employed on the fraud insurance database by Decision Tree and Bayesian algorithm for identification. Other like Nearest Neighbourhood established on pruning rules [5] and association rules has been employed on automobile insurance dataset for building of training model and evaluate the efficiency [6-7]. However, various methods have come in to light for developing an efficient automobile insurance fraud identification system, although, nearly all systems used earlier depicts more deflection in relation to accuracy as these systems require mostly all attributes that exists in the automobile fraud insurance database.

Gopikrishna Panda[1], Sunil Kumar Dhal[1], Rabinarayan Satpathy[1], Subhendu Kumar Pani[2]

The remaining part of the paper is arranged as follows: Section 2 comprises the research objective related insurance fraud in automobile. Section 3 explains working of Attribute Extrication using Discrete Wavelet Transform. Section 4 evaluates methods which are utilized for feature selection by Principal Component Analysis elaborately. Section 5 explains method of identification by the use of Spiking Neural Network (SNN) and Normalised Approximate Descent (NORMAD) algorithm. The proposed technique is elaborated or explained in details properly using flow diagram. The results and discussions inferred from the experiments are evaluated and then required discussion were made in section 6. Finally, section 7 reflects present work-related conclusions and opened a new direction for extension in future.

2. Data preparation for the expert system

Although the key motive of this paper is to coin a algorithm on standard statistical learning for fraud identification well apt mainly for any highly asymmetrical information from an insurance company for research work, we procure simple and quite a sample database from Data Preparation for Data Mining book [8] to reflect the well coined algorithm. The database comprises of 15,420 observations which indicates 32 predictor variables and a diverging reaction variable for fraud detection. The data were collected over a three-year period from 1994 to 1996. 30 definite variables, one continuous variable, and detection variable exist in the database. However, the algorithm proposed is precision for scam identification (classification) other than prediction of insurance premium, neither were time-sensitive variable. Every definite variable was converted into simpleton variables. The binary reaction variable explains in case the claim was classified as false or true. Out of 15420 observations, 923 (6.4%) claims sorted as scam within the database. In the database, there is no missing value vectors. The information is generally unbalanced in insurance fraud identification and it is demanding to create a classification structure with such a unbalanced database.

Nearly 3 years of cases were utilized in this database. The scarcity stage changes that effects these vectors would not be shown in tiny period of 3 years. Although, even since only 3 years of information were utilized in the selected database, the proposed algorithm does not relay on the years of data made due to proposed algorithm takes the years as an ordinal variable. Our algorithm can be utilized to current insurance data made for an escalated number of years with observations in million practically.

The identifier variables include various population related variables like age, gender, marital status, etc. Various variables explain the automobile implicated in the claim such as type, make, price, age of vehicle, etc. Further variables explain the claim like time of year, filing of police report, witness present, etc. The remaining of the variables explain the kind of insurance policy such as deductible, policy type, etc. The variables are concluded in Table I.

Table I. Automobile claims data Details.

Variable	Description
Month	From January to December
Week of Month	1, 2, 3, 4, 5
Day of Week	From Sunday to Saturday
Make	Accura, BMW, Chevrolet, Dodge, Ferrari, Ford, Honda, Jaguar, Lexus, Mazda, Mercedes, Mercury, Nisson, Pontiac, Porche, Saab, Saturn, Toyota, VW
Accident Area	Rural, urban
Day of Week Claimed	From Sunday to Saturday
Week of Month Claimed	1, 2, 3, 4, 5
Month Claimed	From January to December
Sex	Male, female
Marital Status	Divorced, married, single, widow
Age (Continuous)	Ages range from 16 to 80
Fault	Policyholder, third party
Policy Type	Sedan—all perils, sedan—collision, sedan—liability, sport—all perils, sport—collision, sport—liability, utility—all perils, utility—collision, utility—liability
Vehicle Category	Sedan, sport, utility
Vehicle Price	(Less than $20,000), ($20,000 - $29,000), ($30,000 - $39,000), ($40,000 - $59,000), ($60,000 - $69,000), (greater than $69,000)
Policy Number ID	variable
Rep Number	1 - 16
Deductible	300, 400, 500, 700
Driver Rating	1, 2, 3, 4
Days Policy Claims	15 - 30, 8 - 15, more than 30, none
Days Policy Accident	1 - 7, 15 - 30, 8 - 15, more than 30, none
Past Number of Claims	1, 2 - 4, more than 4, none
Age of Vehicle	2 years, 3 years, 4 years, 5 years, 6 years, 7 years, more than 7 years, new

Age of Policy Holder	16 - 17, 18 - 20, 21 - 25, 26 - 30, 31 - 35, 36 - 40, 41 - 50, 51 - 65, over 65
Police Report Filed	Yes, No
Witness Present	Yes, No
Agent Type	External, internal
Number of Supplements	1 - 2, 3 - 5, more than 5, none
Address Change Claim	Under 6 months, 1 year, 2 - 3 years, 4 - 8 years, no change
Number of Cars	1, 2, 3 - 4, 5 - 8, more than 8
Year	1994, 1995, 1996
Base Policy	All perils, collision, liability
Fraud Found	0, 1 (response)

The foremost variable shading of the data, Policy Number (the detection variable) was removed because it represents no meaning to the process. Collinearity in many folds within the predictor variables was evaluated by the variance inflation factor (VIF). When the VIF for a variable was more than 10, then that variable was meant to be as highly corresponding with other predictor variables and was removed from further evaluation. The succeeding variables were subsequently deducted from hypothesis based on their VIF: Base Policy, Vehicle Category, Age of Policy- Holder, Month, and Address Change Claim. Thus, there were 26 rest variables available for respective evaluations. These variables were explained as the initial 26 variables to be studied. There were nil observations removed from the database.

From the original database, learning set and testing set were created. In the paper, learning set was utilized to create total systems. The test set was utilized to examine and process the final outcomes of all the system. Although the whole dataset was steadily unbalanced with 14,927 non-scam cases and 923 scam cases, the learning set was built to provide steadiness to the data for more accurate outcomes. The learning set was inconstantly chosen for 1000 observations by veneered uncertain sampling. Five hundred of 1000 observations were inconstantly chosen from the 14,497 non-fraud cases and the rest 500 observations were inconstantly chosen from the 923 scam cases. The test set included the remaining of 13,997 non-scam cases and 423 scam cases, so therefore the smaller dimension of the test set was 14,420.

3. Discrete Wavelet Transform

As the traditional method of Fourier transform is used very often in the analysis of the insurance data it proved to be less efficient due to its trade off among temporal resolution and frequency resolution. An alternative solution to this problem is wavelet transform, which has been comparatively current advancement in area of digital signal processing by [9], though it has been found to have been invented separately in diverse fields of mathematics, quantum analysis and in electrical engineering [10]. Application of wavelets has been in various domains, such as time series data compression, filtering of noise from data and detection of features [11]. The representation of the signal in the Fourier transform is delineated into a fundamental wave of sine and cosine. The wavelet transform also utilizes a logic, that is elements are defined scale-invariant, as is clearly understood that the basis seems to be the same at all scales, and the basis is space localized. The outcome is that in the wavelet representation, the signal at separate resolutions seen in different window sizes as it can be visualized just as a building and its windows at the same instant of time. On a large scale, the group of buildings can be seen and this can be viewed to get the global features. To look for the window of the building, closer focus is needed and to get local features. A closer look can be made to view hooks on the window. Different scales can be used to view all groups of buildings, building, window and even hook on the window. The major dissimilarity between Fourier and wavelet analysis is that flexible size of window is sufficient for wide spectrum stationary signals such as database (for low frequencies, large windows are used and for higher frequencies, small windows are used). Mother wavelets Ψ_m referred to as the basic feature and various choices to be obtained experimentally for the particular application. For instances of some mother wavelets include the simplest Haar wavelet, that is discontinuous step function. One of the disadvantages of discontinuity in a few domains like audio data, video data or data matrix is not suitable, whereas its advantage lies in random transitions like the failure of the machine [12]. The apt wavelets are the Daubechies wavelets (dbN) reflects on fact that the evenness of the wavelets rises as N rises (db) for the database. To build other parts of the wavelet standards, the mother wavelet is increased and converted by factors i and j using:

$$\psi_{i,j}(t) = \frac{1}{\sqrt{|i|}} \psi_m \left(\frac{t-j}{i} \right) \tag{1}$$

The measure of stretching or compression of the mother wavelet is based on parameter $i \neq 0$ (depending on whether i is greater than or less than 1). Thus, high-frequency components that are introduced to the wavelet family, as i is small as a result of wavelets, can capture high frequencies of the signals. Similarly, to get low-frequency signals, the introduction of the low-frequency component to the family of the wavelet is done. The amount of shifting of the wavelet along the horizontal axis is determined by parameter j. If j>1 that makes wavelet shifts to the right then shifting it to the left with j<1. As a result, the onset of that wavelet specifies parameter j. Subsequently, the

scaled wavelet is defined as daughter wavelets whereas main wavelets are called the wavelet function (mother wavelet) and function of scaling (also called the father wavelet).
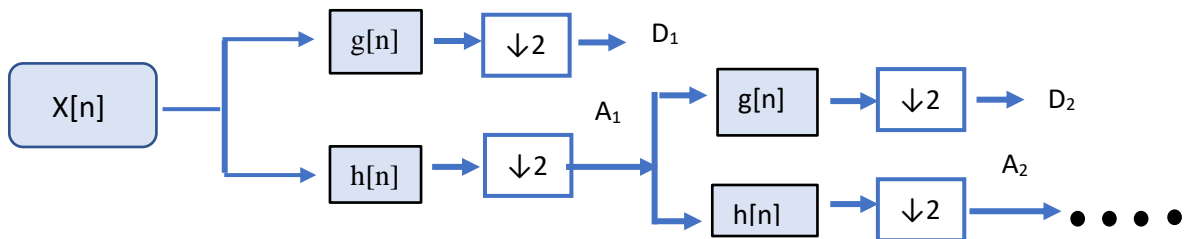
Wavelet Packet Decomposition

Wavelet-based features are also used in fraud event detection in [13]. First, wavelet packet decomposition trees of each signal are derived. Then, the features such as spectral centroid, sparsity, node energy, and spectral spread are extracted from the child nodes of the wavelet tree. The application of wavelets to the digital signal gives rise to separate the data into an approximation (high frequency) part and a detail (low frequency) part of the signal into the matrix. Due to this, wavelets can be used as low-pass and high-pass filters. Analyses of this filtered segment can be performed by wavelet again with the scale with shorter value typically half of the scale giving rise to daughter wavelet. Usually, the approximation parts of the signal contain actual information, that's why this part has to be analyzed again instead of both the detail part and approximation part(coefficients). But this cannot take place separately, so another technique can be used like the wavelet packet decomposition [13]. This method produces a tree of wavelet decompositions, where there are M levels at the tree which is again starting with the head at 2M that produces a rich spectral analysis.

The levels of decomposition are measured with the requirement of the application of wavelet packet decomposition. The method is applied to the data matrix for analysis again even after filtering for segmentation. The segmentation and feature extraction of the signal after filtering data can be achieved by this approach to get the approximation part at each section. The experiment or any convention through which can be estimated is by using information theory. The information contained in the approximation part of the signal as well as the detail parts contains noise that can be removed. In the area of information theory, the amount of uncertainty or disorder in a system is defined as Shannon entropy [13]. The amount of information retained in the provided signal is due to this Shannon entropy. The concept of entropy for using wavelet to evaluate accuracy at an optimum level for the use of selective wavelet is high. We use this computation at each node to choose whether or not to retain a node and stopped creating the tree up to the point where all of the nodes contained noise were removed by this computation, meaning that the signal was fully described.

4. Principal Component Analysis applied to the proposed Features

In order to reduce the data features, this work uses PCA as described in [14]. PCA is used so that the overall

Figure 1 Discrete Wavelet Transform Sub-Band Decomposition

feature vector can be reduced, and only optimum features are available for classification. This assists in reducing the training delay, and improving the overall accuracy of classification. The following data features are evaluated in this work,

- Wavelet features that represent the wavelet domain data (majorly spatial features)
- Permutation Pair Frequency Matrix (PPFM) used to represent sound in terms of adjacent values of data samples

The following table demonstrates the number of features evaluated for each of the feature extraction techniques,

Feature Vector	Number of features	Components per feature	Total features
Wavelet	8000	4	32000
PPFM	16000	2	32000
		Total	64000

Thus, a total of 64000 features are evaluated for each sound sample. Most of the feature values are repetitive, and can be reduced to a much lower number. Thereby initially a variance calculation is done for these features. The variance is not calculated within the features of the same sound, but across different sounds of the entire training set. The following formula is used to evaluate variance for each of the 64000 samples,

$$V_i = \sqrt{\frac{(f_i - \frac{\sum f}{N_t})^2}{N_t - 1}} \; \dots \tag{2}$$

Where, V_i is the variance of the ith feature (i goes from 1 to 64000), f_i is the value of the feature, and N_t are the total number of feature values for the given feature in the training set. Once the variance for each of the 64000 features is obtained, then eigen values are obtained for these vectors. The variance is plotted on a XY axis, which X axis being the feature number, and Y axis being the variance of the feature. For simplicity, consider that there are 6 features, for which the variance is plotted as shown in the following figure,
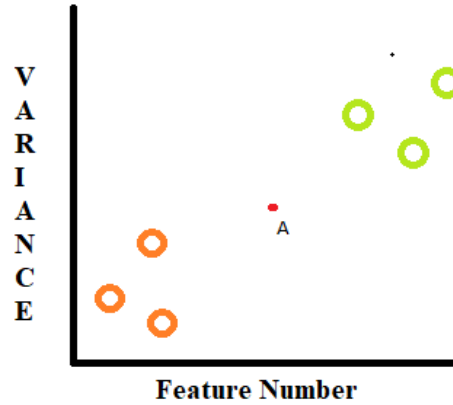
Figure 2 Sample 6 features plotted against variance

The value 'A' is the mean between these points. Now, we evaluate the best fit line between these points, which aims to reduce the distance between these points, and showcase the line as follows,
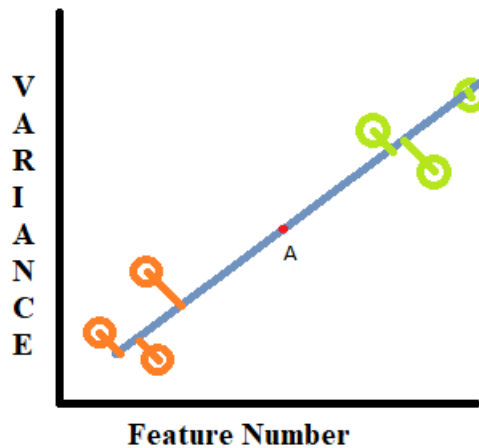

Figure 3 Best fit line between these points

The axis is shifted on the point A, so that the point A and the origin of the axis are coinciding with each other. This helps in evaluating the Eigen values,
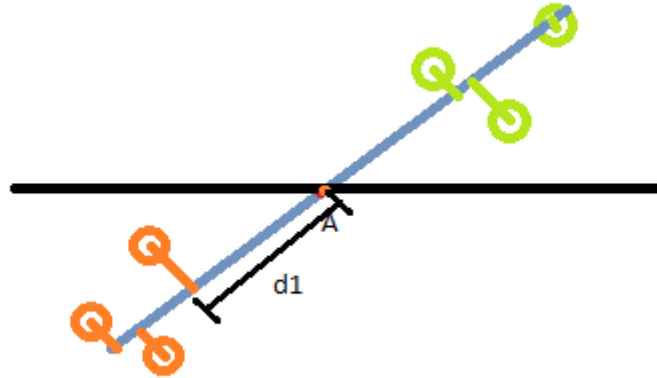
Figure 4 Axis coinciding with the mean value

Find the distances d1, d2, d3, d4, d5 and d6 using Pythagoras theorem, and mark the values d1 to d6 as the eigen values of the features. Now shift the axis making it orthogonal to d1, and evaluate the other feature vectors from d11, d12, d13, d14, d15 and d16, as shown in the following figure (the red line is the orthogonal line),
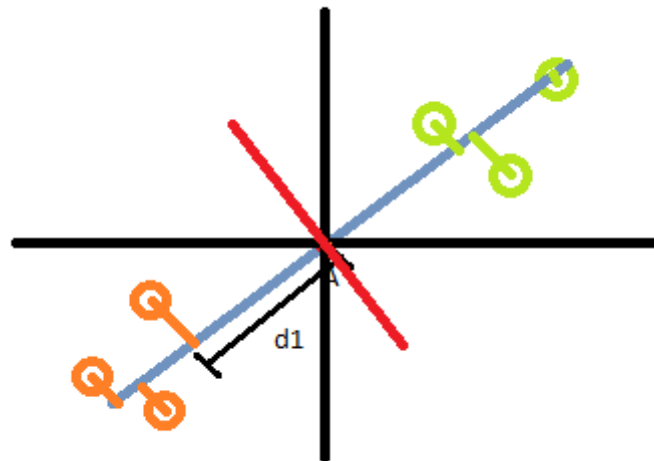

Figure 5. Orthogonal axis to the feature vector d1

Similarly, the axis is made orthogonal to each of the features, and the following matrix is evaluated,

$$\begin{matrix} d1 & d11 & dn1 \\ d2 & d12 & dn2 \\ dn & d1n & dnn \end{matrix}$$

The matrix basically consists of all the eigen values (or principal components) of the features. A singular value decomposition (SVD) is applied to these features, and a single decomposed value is found for the given matrix.

$$SVD = \sum d_i * \sigma_i * f_i \dots \tag{3}$$

Where, d_i is the eigen value, σ_i is the variance of the feature, and f_i is the feature vector value. All the positive values from these SVD values are considered for feature evaluation, while negative or zero values are removed. For our training set, the total number of features got reduced from 64000 to approximately 8000 when using PCA, thereby improving the system speed and accuracy of classification.

4. Spiking Neural Network

A spiking neural network (SNN) is a neurocomputing recognition method that is motivated by how the human brain works with the information. In literature, SNN is established as " enormously parallel interrelated networks of basic (usually adaptive) elements and their heterarchical establishments which are deliberately to interchange with the elements of the real world in the similar way as nerve system in human brains do". Human brain is presumed to be comprised of billions of interrelated neurons of many layers. Human brain neurons have capability to learn data. Due to this reason, humans are surprisingly efficient at analyzing the world that they visualize. A simple example of the task is Handwriting recognition. Adapted

by various other handwritings through many years, an average human brain is ability of apprehend the handwritings of different people promptly. For practical use, it is difficult to employ human brain activity of understanding letters and characters into a program to make model. Even though, it is easier to think then to apply and employ model. The cause is that the differences in handwritings of various people makes it really rigid to detect accurate models. This results in majorly lesser success rate for computers than humans.

SNNs point of view to this issue is in a same manner with human brain. It directs each attribute (or, basically, training vectors) as input, towards an input layer, an output layer and (optional) hidden layer(s) of artificial spiking neurons and begins to adapt the network with each training vectors. A cost function is selected to determine the error between the desired output and the estimated output. The task of training is to minimize this cost function iteratively.

Let's discuss the basics of spiking neural network. The key components of SNN are Spiking Neurons and the synapses that interconnect them. As we can recall from biology, the unit of nervous system is nerve cell which transmits information by passing action potential to another neuron connected to it. So, a spiking neuron in fact spiking artificial neuron does incorporate this key aspect while modelling this behaviour.

figure 6 Neuronal integration and Spike Communication
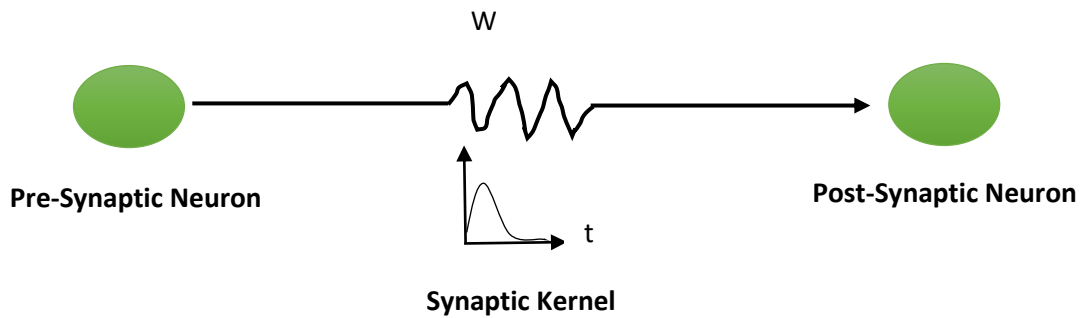
Consider here a system of two neuron input neuron (Presynaptic neuron) and output neuron (postsynaptic neuron) both are interconnected by synapses having strength w as shown in figure 6. This synapse is modelled with double decay exponential kernel. These are some mathematical modelling features that people incorporate. To understand this system, a presynaptic neuron issues a stream of spikes which then get translated to postsynaptic current when spikes are passed to synaptic kernel and this post synaptic neuron then integrate the incoming current that is reflected in the membrane potential which rises as per the incoming current. So as when the membrane potential of the neuron exceeds the threshold potential then it issues spikes.

The neurons used in my work is simple leaky integrate and fire neurons. The LIF neuron captures the key aspect of integrating the incoming current and issuing spikes whenever the threshold is exceeded.

That is well described by the differentiation equation of membrane potential.

$$C_m \frac{dV(t)}{dt} = -g_{lm}(V(t) - E_{lm}) + I_{sy}(t) \tag{4}$$

Whenever $V(t)$ exceeds E_{lm}, spikes is issued and $V(t) = E_{lm}$ it is reset.

The simulation in this work uses, $C_m = 300pf$ is the capacitance of the membrane, $g_{lm} = 30ns$ is the leak conductance, threshold voltage $V_T = 20mv$ and I_{sy} is integrated synaptic current input to the neuron. An SNN prototype is given in Figure 7. The membrane potential is reset to its stable position $E_{lm} = -70mv$ after the issue of spike. When the voltage $V(t)$ is greater than V_{TH}, generation of spike occurs and is transmitted to downstream synapses. The current potential $V(t)$ is equal to E_{lm} remains in it for small period after issue of spike. The small period $Trf = 3ms$ where next spike is not issued. The leaky integrate and fire model acts like a nonlinear spatial filter with $w_1, w2, w3, w4 \ldots w_n$ as synaptic weight as shown in figure 5.
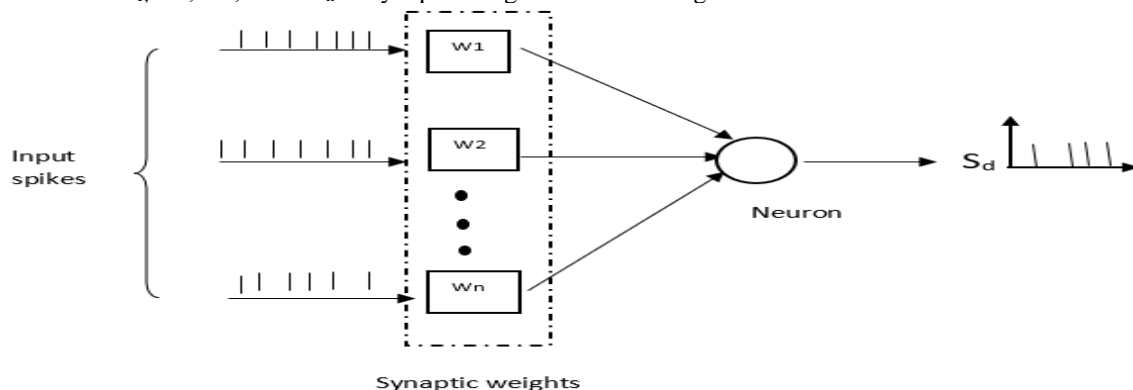
Figure 7 SNN prototype

The model of the synaptic current kernel $k_i(t)$ is a framework of the variables such as rising time constant $\tau_1 = 5$ ms and decay time constant $\tau_2 = 1.25$ ms, respectively. The input received by neuron is from n synapses and time of spike arrival at the jth synapses is denoted as t_j^1, t_j^2, $t_j^3 \ldots t_j^{nj}$. Then the input at the jth synapse is converted in to post synaptic current I_{sy} is given by following equation

$$k(t) = \sum_n \delta(t - t^n) * \left(e^{-\frac{t}{\tau_1}} - e^{-\frac{t}{\tau_2}} \right) \tag{5}$$

where $k(t)$ is synaptic kernel

$$I_{sy}(t) = w \times k(t) \tag{6}$$

As the neuronal integration is highly nonlinear due to abrupt resetting of membrane potential. Synaptic Kernel is modelled by double decay exponential kernel and that when it is weighted with factor w give rise to post synaptic current.

Let me discuss the learning algorithm that is the spike based supervised learning as called as NormAD (Normalised Approximate descent). So the meaning of supervised learning in SNN is to make the neuron issues spikes at desired instant of time at given a set of input signal.

And to do so we consider a parameter that is typical synaptic weight w so that neuron issues spikes at utter desired instant of time.

As part of learning Algorithm, we define this error function as the difference between desired spikes and observed spikes that is the one issued by the neuron. That is used to feedback term to update Synaptic weight.

$$\text{error }(t) = S_d(t) - S_o(t) \tag{7}$$

$$J(ws) = \frac{1}{2} \int_0^T |\text{error}(t)| \, (V_{desired}(t) - V(ws, t))^2 dt \tag{8}$$

In order to do that we define a cost function as w. r. t. to synaptic weight w as the integrated difference between the desired to the observed membrane potential.

And applying gradient descent rule, that is typically used is as optimization of neural network we can find the instantaneous weight update term as derivative of the cost function which can be written in term of difference of derivative of the membrane potential.

$$\Delta w(t) = \eta_r(t) \nabla_w J(ws, t) \tag{9}$$

$$\Delta w(t) = \eta_r(t) |\text{error}(t)| (V_{desired}(t) - V(ws, t)) \nabla_w V(ws, t) \tag{10}$$

$$\Delta w = z \int_0^T \text{error}(t) \frac{\hat{d}(t)}{\|\hat{d}(t)\|} dt \tag{11}$$

Where $\hat{d}(t) = k(t) * \hat{h}(t)$,

$\hat{h}(t) = \exp(-t/\tau_L) \, u_s(t)$

k(t) represented as synaptic kernel

$u_s(t)$ is the Heaviside step function

τ_L = leak time constant of the membrane=10ms

Now the membrane potential as I discussed in my previous slide is represented by difference equation. Membrane potential equation is highly nonlinear due to spike occurring whenever V(t) exceeds $E_{lm.}$ So, considering the time interval between spikes we can solve that differential equation in closed form manner so that we obtained an expression we clearly see the dependency of synaptic weight on membrane potential in this manner.

So, it is easier to calculate the derivative of membrane potential and also on further approximation that is by reducing the leak time constant τ_L of membrane potential. Basically, what we do to make spike issue by neuron that is spike coming to the neuron sparse so that there is no dependence of derivative term on the synaptic weight. There is further Normalization done on the voltage derivative term so that this dependence on $V_{desired}$ is completely eliminated.

All we know about supervised learning task is the time instant of desired spike $S_{desired}$ term we have no knowledge of $V_{desired}$ term. So finally, our expression of weight update is the term depended on error and normalised voltage derivative term. Hence, this gives us the closed form expression of synaptic weight using the incoming desired spike trains.

6. Simulation results

In this section of the paper, there is a detailed description of how the proposed technique is simulated. For simulation, the experiments are implemented in MATLAB 2015 b software. The test platform is Intel core 3i 8th generation, 2.2-GHz CPU, 4-GB RAM processor with Windows 7 operating system. As illustrated in figure 8, the training data is taken as the input which is the first module of the system, which is then given to the feature extraction module after pre-processing.
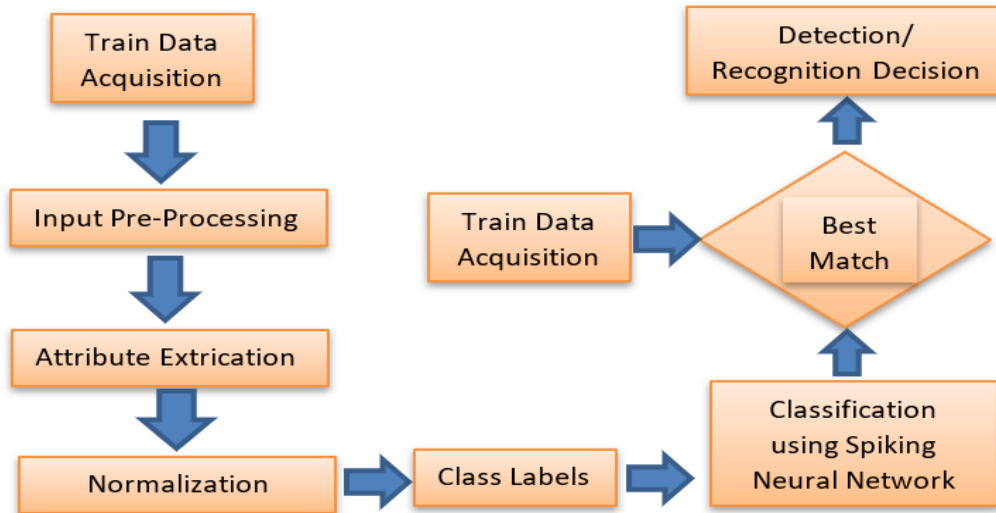
Figure 8 Flowchart for the design of automobile Insurance fraud detection system

Extrication of Attributes: Attribute extrication techniques used are the Discrete Wavelet Transform (DWT) and permutation pair frequency matrix (PPFM) in this proposed system. Ultimately in DWT, the approximate coefficient of attribute vector is 40 in dimension with the usage of Daubechies 4 and decomposition level of 10 which gives a optimum result. In permutation pair frequency matrix PPFM with classifier is efficient due to the permutation window is 5 samples with a time lag as 1 just as in [15-16]. Delay time taken is more in SNN with permutation pair frequency than in SNN with wavelet

For classification purposes, the Spiking Neural Network (SNN) and Artificial Neural Network (ANN) Model are used to classify the system. The system is divided into two main models, first is the training model where the data are input and attributes are calculated and stored in the database which can be used later to generate prototype vectors for specific data. Then the ANN model is included after attribute extrication for clustering vectors together in the feature space. This ANN model is only included in the training phase of the proposed technique. Once the system is trained it is then tested by implementing the same techniques.

The system begins by taking the insurance data as input and feature extraction was calculated by different methods like Wavelet and Permutation Pair Frequency Matrix (PPFM), which are classified by using SNN & ANN with the features available in the database. A traditional method like the attribute extrication method along with ANN is used for classification comparison.

The following algorithm steps are used for the development of the classification

1. The data was applied for feature extraction to the wavelet-based and PCA algorithm.

2. The features are given to the PPF generation matrix, wherein feature variation was evaluated and distinctly identifiable matrices are generated.

3. Matrices are found for every instance and applied to the SNN training layer.

4. The performance parameters of SNN are evaluated, and upon satisfactory performance, the SNN configuration was finalized.

5. For any new data, steps 1 to 3 are repeated and the DWT matrix was applied to the trained SNN.

6. The obtained class was evaluated and system accuracy was checked.

7. If the accuracy was lower than expected, then the SNN configurations are modified, and the process is repeated.

Evaluation of Proposed Algorithm

In order to evaluate the proposed algorithm, we decided to test the entire dataset on the algorithm. And then evaluate the results in terms of both accuracy and computation time. The following formulas were used,

$$\text{Accuracy} = \frac{\text{Number of sounds correctly classified}}{\text{Total number of sounds tested for that class}} \quad (12)$$

$$\text{Computation Time} = \text{Completion time stamp} - \text{Start time stamp} \quad (13)$$

Where, time stamp is the standard Unix epoch timestamp which is the number of seconds since 1st January 1970
Based on these evaluations, the results were evaluated which are mentioned below.

Gopikrishna Panda[1], Sunil Kumar Dhal[1], Rabinarayan Satpathy[1], Subhendu Kumar Pani[2]

| No of epoch	ANN with PPFM		ANN with Wavelet		SNN with PPFM		SNN with Wavelet	
	Computation Time in Sec	Accuracy%	Computation Time in Sec	Accuracy%	Computation Time in Sec	Accuracy%	Computation Time in Sec	Accuracy%
10	0.92	75	0.82	78	0.71	88	0.52	94
20	0.92	75	0.85	78	0.72	86	0.52	94
50	0.95	75	0.88	75	0.75	85	0.54	92
100	0.98	72	0.83	75	0.81	85	0.60	92
200	0.99	71	0.85	75	0.83	87	0.61	93
500	0.94	70	0.86	73	0.84	85	0.63	92
1000	0.94	70	0.86	73	0.84	85	0.67	92
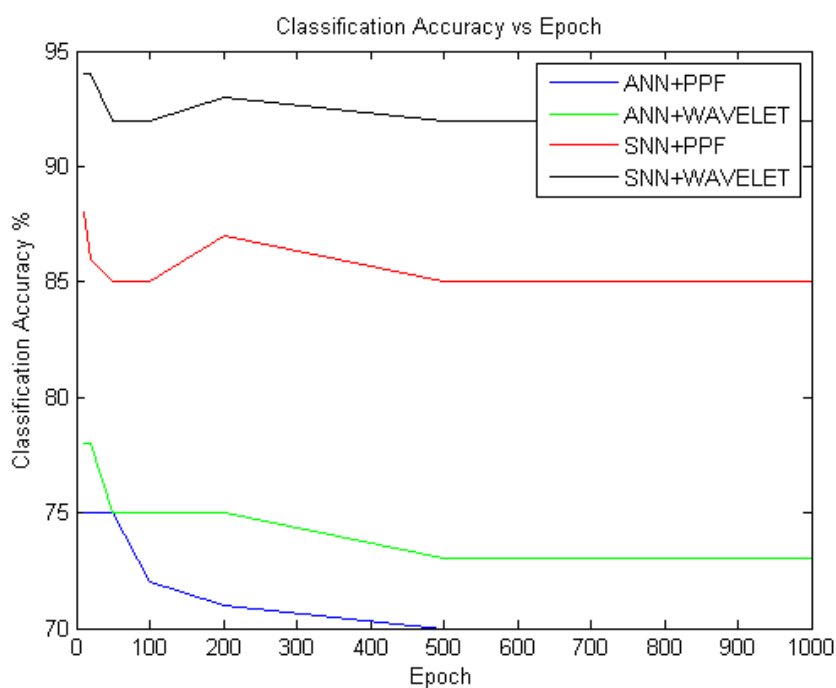
Figure 10 Graphical representation of GMM and SNN Results

Figure 10 Graphical representation of GMM and SNN Computation Time
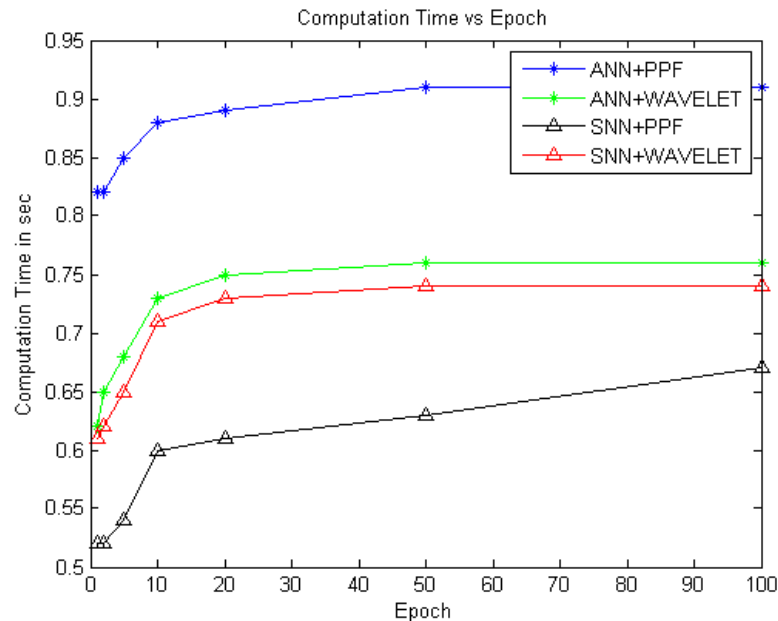
With the computation time delay between 0.6 to 0.8s and the classification accuracy escalated from 92 to 94 % using SNN along with wavelet is depicted in the graphical representation in figure 9 & figure 10. In figure 9, the black line depicts growth of classification accuracy of SNN along with wavelet that is from 92 to 94% depending upon the number of epochs. The black line appears to decline to any extent (epoch 5 to epoch 45) as the epoch increases and then seems a straight horizontal line after 50 to 100 epochs. This black line represented here is analysed as better than all other technique utilized in the level of classification accuracy.

In the case of figure 10, the computation time delay of SNN along with wavelet is between ANN along with Wavelet and SNN along with permutation pair frequency. Though in SNN along with wavelet is best in classification accuracy but the computation time delay of SNN along with Wavelet is moderate in performance, therefore it is better than all other three techniques performed here.

7. Conclusion

This paper focusses on the competitive identification issue of automobile insurance fraud were analysed and talked through. Owing to the issue situation, comprehensive and extensive analysis and extraction are executed through data mining and analysis, and two bands of attribute extrication are carried out depended on the traditional attribute extrication mode. Focussing at the issue of imbalanced category distribution in the automobile insurance fraud recognition situation, the SNN model with NormAD algorithm was coined. This algorithm was utilized to resolve the issues of inadequate sample usage, easy overfitting, and low classification rate in the class distribution problem. Ultimately, through the extensive study and experimental evaluation reflected in this article, the results demonstrated that SNN model is the best for now in comparison to other conventional methods.

In the days to comes, we intend to improvise results by using the adaptive attribute ranking semantic algorithm based on natural language understanding (NLP) to enhance the problem of attribute importance screening and analysis.

Reference

1. Yibo Wang, Wei Xu, "Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud,ˮ Accepted Manuscript, 2018
2. Sharmila Subudhi, Suvasini Panigrahi, "Use of optimized Fuzzy CMeans Clustering and Supervised Classifier for Automobile Insurance Fraud Detection," Elsevier, 2017
3. Luis Alexandre Rodrigues and Nizam Omar, "Auto Claim Fraud Detection Using Multi Classifier System,ˮ pp. 37–44, 2014.
4. [Leila Goleiji, M. J Tarokh, "Identification of Influential Features and Fraud Detection in the Insurance Industry using the Data Mining Techniques,ˮ Majlesi Journal of Multimedia Processing, Vol. 4, No. 3, September 2015.

5. Yaqi Li, Chun Yan, Wei Liu, Maozhen Li, "A Principle Component Analysis-based Random Forest with the Potential nearest Neighbor method for Automobile Insurance Fraud Detection Identification," Journal of applied soft computing, July 14, 2017.

6. Yaqi Li, Chun Yan, Wei Liu, *"Research and Application of Random Forest Model in Mining Automobile Insurance Fraud," IEEE, 2016.*

7. *G. Ganesh Sundarkumar, Vadlamani Ravi, V.Siddeshwar, "A Novel hybrid Undersampling Method for Mining Unbalanced Datasets in Banking,"Elsevier, 2015.*

8. *Pyle, D. (1999) Data Preparation for Data Mining. Morgan Kaufmann Publishers, Inc., San Francisco.*

9. Alfred Mertins, "Wavelet Transform. In Signal Analysis: Wavelets, Filter Banks, Time Frequency Transforms and Applications," chapter 8, pp 210–264. John Wiley & Sons, 2001.

10. Amara Graps, "An introduction to wavelets," IEEE computational science and engineering, vol. 2, no. 2, pp 50–61, 1995

11. S. Patil, G. S. Chandel, and R. Gupta, "Performance analysis of steganography based on 5-wavelet families by 4 levels-DWT," International Journal of Computer Science and Network Security (IJCSNS), vol. 14, no. 12, pp 56–61, 2014.

12. C. S. Burrus, R. A. Gopinath, and H. Guo, "Generalizations of the basic multiresolution wavelet system," In Introduction to Wavelets and Wavelet Transforms: A Primer, pp 98–145, Prentice-Hall, Inc., 1997.

13. H. F. Ma, J. W. Dang, and X. Liu, "Research of the optimal wavelet selection on entropy function. In Future Control and Automation," pp 35–42. Springer, 2012. doi: http://dx.doi.org/10.1007/978-3-642-31003-4 5.

14. I. Potamitis, S. Ntalampiras, O. Jahn, and K. Riede, "Automatic bird sound detection in long real-field recordings: Applications and tools," Applied Acoustics, vol. 80, pp 1–9, 2014. doi: http://dx.doi.org/10.1016/j.apacoust.2014.01.001.

15. R. Mohanty, B. K. Mallik, S. S. Solanki, "Automatic bird species recognition system using neural network based on spike," *Applied Acoustics*, vol 161, pp 1-8, 2020.DOI: 10.1016/j.apacoust.2019.107177.

16. R. Mohanty, B. K. Mallik, S. S. Solanki," Normalized approximate descent used for Spike based Automatic Bird Species Recognition System, Springer-International Journal of Speech Technology, pp 1-10, 2020. DOI: 10.1007/s10772-020-09735-6.