ISSN 2347 - 3983
Volume 8. No. 5, May 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter33852020.pdf
https://doi.org/10.30534/ijeter/2020/33852020

Decision Support System for Detection of False Agricultural Insurance Claims using Genetic Support Vector Machines

Rajesh Budihal[1], Dr G Komarswamy[2]
[1] Research Scholar, Department of Computer Science and Engineering, Jain Deemed to be University, Bengaluru, India
[2] Associate Professor, Department of Computer Science and Engineering, Jain Deemed to be University, Bengaluru, India

ABSTRACT

Insurance process if highly indeed for agricultural sector. Agriculture is purely a climate-based process and the climate or the nature is in the control of anybody. Hence, without insurance process and if anything, wrong happened with the farmer, then the farmer has to bear a heavy financial loss. So, to stabilize the investment, income and life of farmer the insurance process is an unavoidable mechanism. The proposed scheme Genetic Support Vector Machines (GSVM) is based on a study which has been done in the year of 2017-18 for Karnataka state. There are four districts like Dharwad, Haveri and Vijayapura, and Belgavi have been included for this study. In this paper we have proposed a machine learning-based decision support system which validates the genuineness of claims made by the farmers.

Key words: Insurance; Agriculture; Farmer, Claim; Authenticity

1. INTRODUCTION

In spite of so much development in science and agricultural technologies still, farming is a climate dependent process. All most every year, farmer suicide is the breaking for all newspapers. As per the news published in [1], more than 12000 farmers have committed suicide between the year of 2015-18. The major hindrances for the agricultural processes are cyclones, floods, draught, earthquake etc. Additionally, inadequate management for agricultural sector is another major problem for which insurance process if the only solution. In the making of disasters human efforts also pays a major role like, selling of fabricated seeds, fertilizers and pesticides etc. [2]. Either it is natural or man-made calamities, but the loss due to this so high which is unbearable by the farmers and that compels them for the suicidal decisions (Ref. Fig 1). In the process of supporting to the agricultural process various policies and acts have been launched by the both state and central government like Krishi Udan, National Agriculture Market (eNAM) and Small Farmers Agribusiness Consortium (SFAC), Pradhan Mantri Krishi Sinchai Yojana (PMKSY), Paramparagat Krishi Vikas Yojana (PKVY), Pradhan Mantri Fasal Bima Yojana (PMFBY), Gramin Bhandaran Yojna, Livestock insurance Scheme, Scheme on Fisheries Training and Extension, National Scheme on Welfare of Fishermen, Micro Irrigation Fund (MIF), and National Mission For Sustainable Agriculture (NMSA) [3].

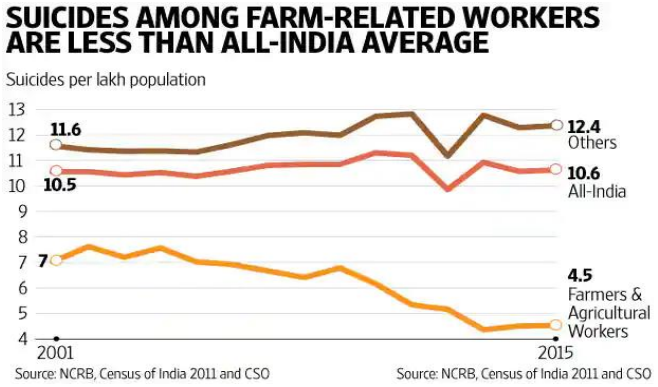

Figure 1: NCRB report on the reasons for farm suicides

The thought of agricultural insurance has been started since 1972 [4]. Though these policies have been existed since a long time but due to non-transparency, partial decision makings, non-payment of loans and high premium rate leads to a catastrophic situation. Again, the revised version of this has been lunched in the year of 1985 where homogeneous area approach has been adopted [5]. In successor to this many such agricultural policies have been introduced to support the farmers of India [6].

Table 1: Problem faced by farmers during claiming (n=240) [4]

Sl. no.	Statements	Garrett score	Rank
1	Late in claim clearance	73.53	I
2	Improper compensation	61.51	II
3	Partiality in loss calculation	56.42	III
4	Tedious Job	52.32	IV
5	Lack of advertisement	48.91	V
6	Zero insurance coverage for crop failure	46.95	VI
7	Forceful insurance imposition	42.24	VII
8	Exclusion of important crops	41.60	VIII
9	Remote bank establishment	40.07	IX
10	No loan no insurance policy of bankers	34.40	X

This work considered the published report in [4] for further analysis. In this report the authors have produced a constraint table by the process of interviewing to explicit farmers. These farmers have experience of insurance claiming as per the report. Total 240 samples have been collected and further that processed has been processed through Garrett's Ranking Technique. Table 1 represents the study report on the insurance behavior. We have considered the same table for our analysis model. Table 2 represents the suggestion proposed by the farmers to avail the insurance with less hurdles.

Table 2: Suggestion from the farmers (n=240) [4]

Sl. no.	Statements	Garrett score	Rank
1.	Claim clearance at the beginning of next session	75.70	I
2.	Explicit insurance office at each block	66.40	II
3.	Proper awareness spreading about insurance benefits	54.91	III
4.	More demonstration on crop-cutting	43.91	IV
5.	Inclusion of wide range of crops	43.62	V
6.	Voluntary insurance opting	40.67	VI
7.	Local establishment of insurance office	38.30	VII
8.	All cropper should be included under insurance scheme	36.55	VIII

The said study also reveals the suggestion from farmer's end to enhance the insurance policies for agricultural sectors. Likewise, many theories have been proposed and discussed for agricultural sectors in the past literature. The insurance for agriculture sector is two-fold. In one hand insurance companies are keep on cheating with farmers [7] [8] [9]. On the other hand, the fraudulent insurance claims are on rise [10]. In this paper, we have motivated to identify or to differentiate the fraudulent claims in compared to genuine insurance claims.

2. LITERATURE REVIEW

As earlier said, the total agricultural process is solely dependent upon external factors like weather condition, climate changing, soil quality, mineral level in soil as well as in water, etc. [11]. The chances of disaster are very high in farming in comparison to the other labor sectors of the society. Hence, the number of claimants for insurance proportionally very high in agriculture sectors. Though, we are reading and observing many cheating cases against farmers, same time vice versa is also true up to some extent. In the past literature there are many decision support systems (DSS) and other methodologies have been proposed to smoothen the disaster -insurance cycle. Natural calamities are the primary villains for farmers. Taking this in account an earth-observation (EO)-based DSS have proposed for the accurate loss calculation with proper insurance claiming [12]. The agricultural processes are normally carried out in acres of lands and in such cases if any disaster happens then loss calculation is the most tedious and hectic job. The manual process of loss calculation or monitoring is both time and labor consuming [13]. To bring advancement in this loss calculation process a Geographic Information Systems (GIS) database based interactive DSS have proposed. According to this model, the remote recording, analysis and evaluation process will become more accurate, efficient and consistent [13]. It can be noted that only providing a sound insurance scheme or an efficient DSS is not the solution towards effective farming process. The farming process need to be strengthened by the amalgamation of advanced process. In Andhra Pradesh, eSagu^{TM} system has been used by the 8000 farms of 16 major crops in 100 villages since 2004 [14]. This system suggests the farmers as the collected real-time field images. The suggestion includes the monitoring of fertilizer uses, additional yielding and pesticide spray. Through this model the loss amount or loss rate can be reduced which is a mutual benefit for both the farmer and insurance companies. In the western countries the major damage of crops occurred because of hailstorms. In [15], a forecasting model has been proposed which generates the ground reality and meteorological information which is very much necessary for the effective farming process [16]. Weather is the primary parameters for agricultural process. If weather get worse then technology cannot do anything to stop the disasters. Hence, weather-based forecasting technique is highly required to protect the farms from the upcoming disasters. This is a risk management strategy with which various other factors are deeply related like sustainability, stability in income and insurance coverage [17] [18]. Although many agronomic tools are available but very few such tools are existing which is applicable at the ground level. In [19], a two-fold DSS have proposed by the researchers where in the first stage a diagnosis DSS at working situation have been analyzed then in the second stage use various prototypes have been discussed. In [20], an DSS have been proposed to improve productivity and environmental decision making. The DSS in agriculture help to make effective decisions based on various stages and information flow.

3. GENETIC SUPPORT VECTOR MACHINES (GSVM)

The GSVM concept was first introduced by John Holland in the year of 1960. The genetic algorithm (GA) is the combination of Darwinian survival of the fittest theory with the properties of genetics like crossover, mutation, and inversion [21] [22]. The GS operates through three primary operators such as 1) crossover, 2) mutation, and 3) reproduction. In this process the problem needs to be converted in to genetic form further, a fitness function has to be formed to work on that platform [23]. During the crossover

phase new offspring are produced from the selected parent chromosomes. Mutation is the process which brings diversity in search space. The reproduction is controlled by the crossover and mutation process. SVM is a statistical machine learning technique proposed in 1995 [24]. This SVM technique can be used in several domains like image processing, bioinformatics, binary classification, etc. SVM is an excellent example of supervised learning [25] [26]. SVM presents the distance between boundary and closest dataset. The GA for differentiating fraud and legal can be presented as:

$$X^t Y_i + C = 0, \qquad x \in R^n, b \in R$$

Where, $\mathbf{Y^n}$ is the training dataset, **X** is the normal to the hyperplane, $\mathbf{R^n \in F}$, is the feature space **F**. The purpose of using SVM is to separate the fraudulent claims from the genuine claims by creating a hyperplane between these two. Fig 2 illustrates the role of SVM to generate a hyperplane between two groups. In this SVM trying to create a maximum margin so that legal claims and false claims can be distinguished [27]. There are few cases which lies in the confusion zone where this differentiation is quite difficult.

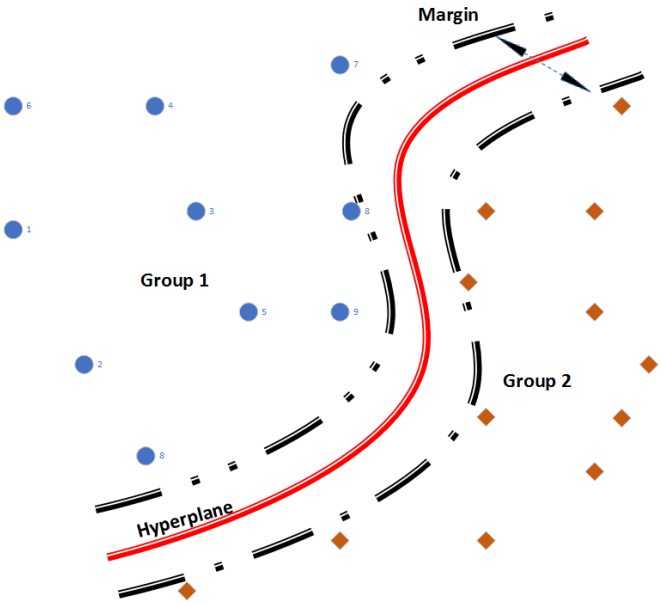

Figure 2: Hyperplane separates the two groups

4. METHODOLOGY FOR GSVM FALSE CLAIM DETECTION

The blueprint of the proposed GSVM model is illustrated in Fig.3. In the blueprint the inputs (i.e. various types of claims) are given to the False Claim Detection Classifier (FCDC). The combine effort of FCDC and False Claim detection model (FCDM) able to segregate the legal and illegal claims. The whole process is the integration of GA, SVM and machine learning techniques [28]. The SVM works through three primary phases such as preprocessing, classification and postprocessing.

4.1. Data Processing

In this work, the required data are collected from the Indian government portal called "KRISHI". Initially, the collected data are found in the raw format. The term raw means, the data are not in uniform format. Many fields are missing or redundant or incomplete hence, there is a need of data preprocessing. For that, data mining tool has been used to bring the data into smooth format so that it can import to next level of process. During data mining, it removes the unnecessary parts of data and brings the complete, non-redundant and smooth data. These smooth data are get stored in excel CSV format further it gets imported to MySQL database. The preprocessing phase includes three phases like filtering, selection and extraction of features and adjustment of features. The collected data is so voluminous, the data mining technique helps to bring the optimized data which relevant and necessary for the feature extraction and for the rest of process. For feature selection WEKA machine learning tool with MATLAB for data processing has been used in this work. Claims are partitioned into two primary categories like 1. Claims with prior approval and 2. Claims with non-approval.

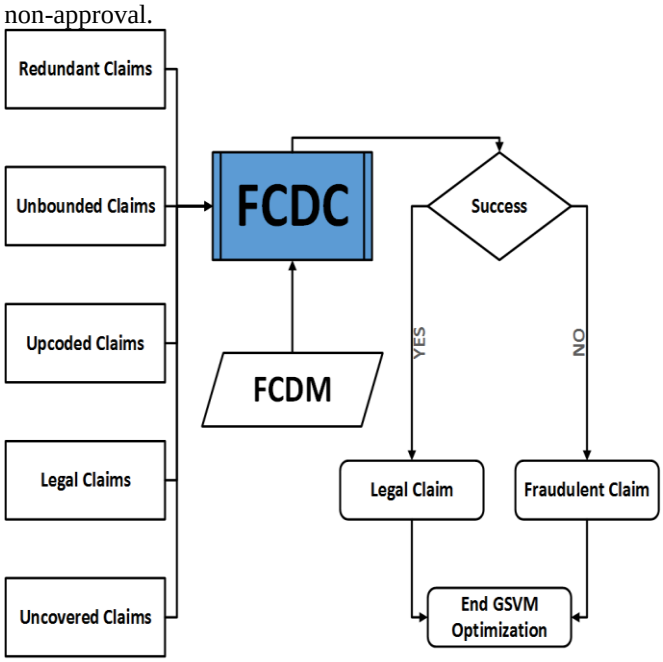

Figure.3: Blueprint of GSVM model

This agricultural insurance process works with few parameters and the same are considered as selected features for GSVM model [29]. The GS based selected features are: cropping date, crop land code, materials used bill, labor bill, and loan approval code. Based on these parameters the optimized formulation is:

$$Min\ (Total_{Cost\,t}) = F(M_{bill}, L_{bill})$$

Subject to:

$$\sum_{i=1}^{n} (M_{i,bill}) \leq A_{loan}, \qquad \forall i, i = 1,2,3, \ldots n$$

$$\sum_{i=1}^{n}\left(L_{j,bill}\right) \leq L_{loan'} \qquad \forall j, j = 1,2,3, \ldots n$$

Where, $\mathbf{M_{bill}}$ is material bill and $\mathbf{L_{bill}}$ is labor bill.

Fig.4 represents the flow-chart of implementation phase of GSVM. This is the combination of FCDM with GA approach. It handles large dataset through kernelization. This flowchart results a matrix which represents the false legal, false illegal, true legal, true illegal and confused bills data [30].

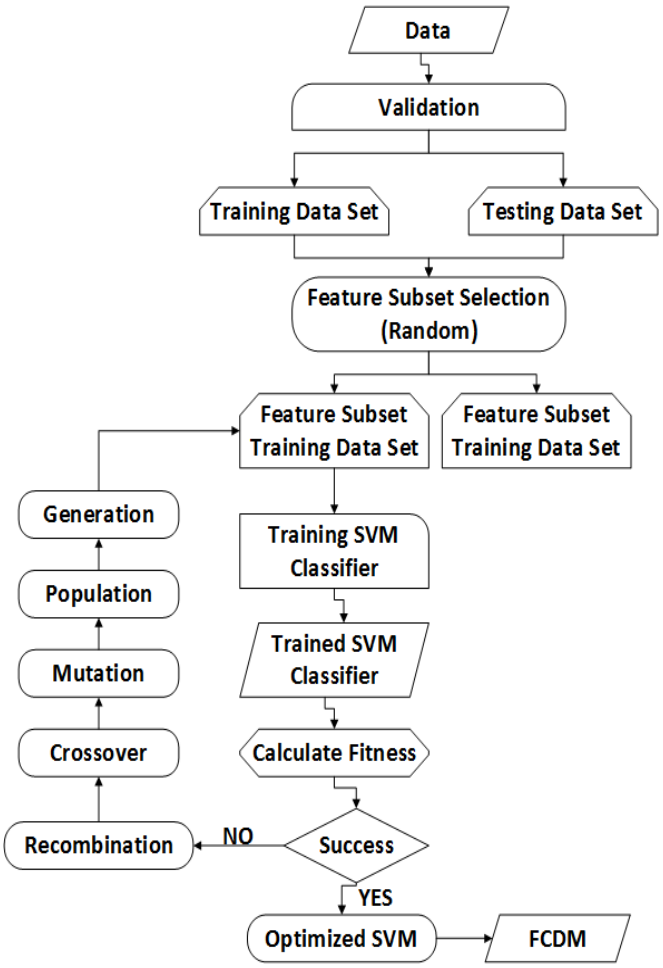

Figure 4: Flowchart of GSVM

The SVM classifier classifies the imported bill based on its features (Ref: Fig.5). The false legal ($\mathbf{F_L}$) claims are actually false but classifier fails to detect it and classify it as legal. The false illegal ($\mathbf{F_I}$) claims are wrongly classified as false where as these are legal claims. The true legal ($\mathbf{T_L}$) claims are legal in true sense. The true illegal ($\mathbf{T_I}$) claims are illegals in true sense. There is one remaining category called confused claims ($\mathbf{C_C}$) which are can't be classified by the classifiers. Hence, the rate of accuracy can be calculated by:

$$Accuracy\ rate\ (A_R) = \frac{T_L + T_F}{T_R}$$

$$A_R = (1 - error) = \frac{T_L + T_I}{T_L + T_I + F_L + F_I} = \mathrm{P}(C)$$

$$Sensitivity = \frac{T_L}{T_L + F_I} = \frac{T_L}{P_L}$$

$$Specificity = \frac{T_I}{T_I + F_L} = \frac{T_I}{N_L}$$

The above formulas calculate the accuracy, sensitivity and specificity of the SVM outcomes.

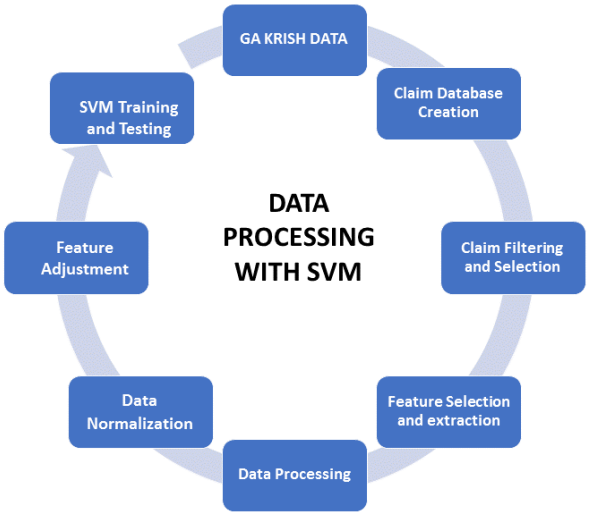

Figure 5: Data processing with SVM

4.2. GSVM implementation

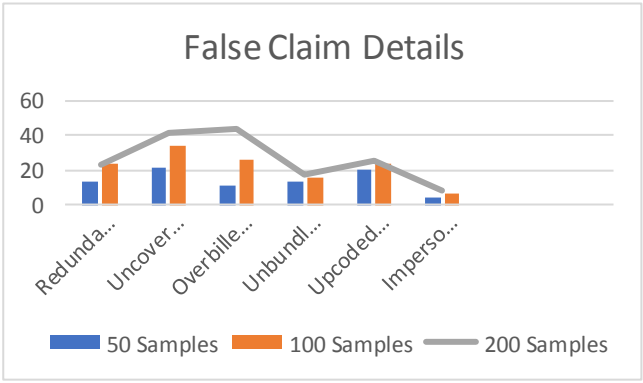

Figure 6: Fraud type distribution

The total GSVM is the combination of four modules such as: 1. MATLAB for computation, 2. Graphical user interface (GUI) for interference, 3. SVM for detection of false claims and 4. Postprocessing of detection. For the simulation purpose we have used three types of sample datasets like dataset of 50, 100 and 200. Fig 6 and Table 3, represents the false claims from the sample data sets. The above formulas calculate the accuracy, sensitivity and specificity of the SVM outcomes. Table 4, shows the performance of linear SVM with different sample data sets.

Table 3: False Claims in 200 Samples

Types of Claims	Samples		
	50	**100**	200
Redundant Claims	14	24	23
Uncovered Claims	21	34	41
Overbilled Claims	11	26	44
Unbundle Claims	13	16	18
Upcoded Claims	20	24	26
Impersonation Claims	4	6	8

Table 4: Performance of Liner SVM with different sample datasets

Data Sets	AR	Specificity	Sensitivity
50	68.42	48.00	76.42
100	70.49	79.11	74.21
200	81.23	82.83	61.49

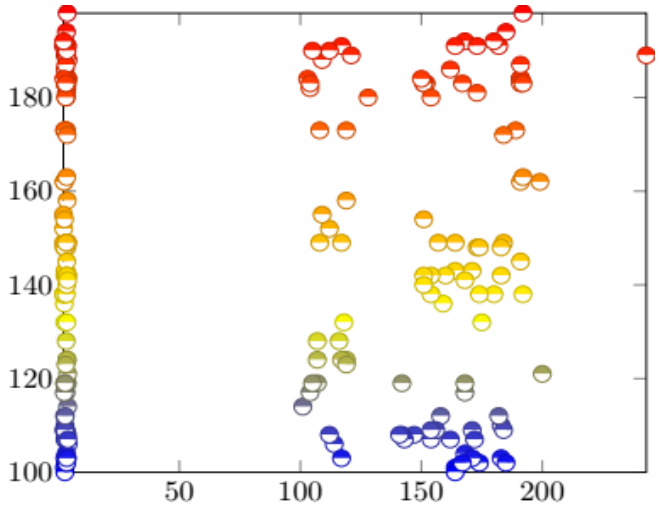

Figure 7: SVM on sample claims datasets

Fig. 7 and Fig. 8, illustrates the five types of claims with four different colors like red color half balls represents FL, orange color half-balls present FI, yellow color balls present the TL and Blue color balls represent the FI and grey color represent the CC. In this simulation the accuracy factors have been analyzed by testing the data. It shoes the performance of GSVM models.

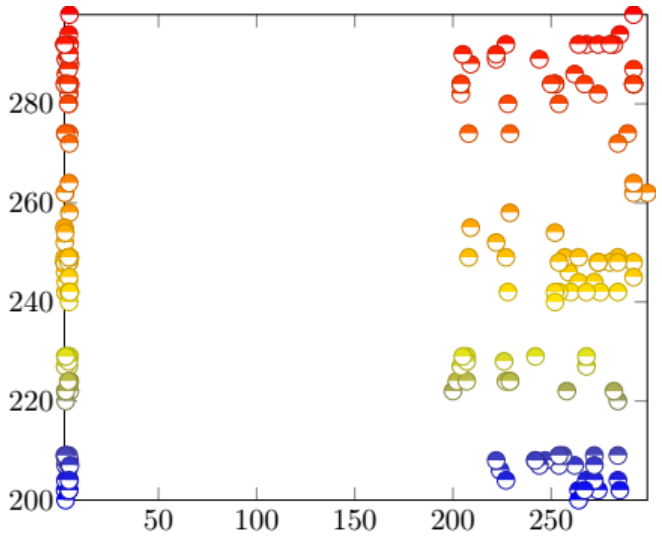

Figure 8: SVM on sample claims datasets

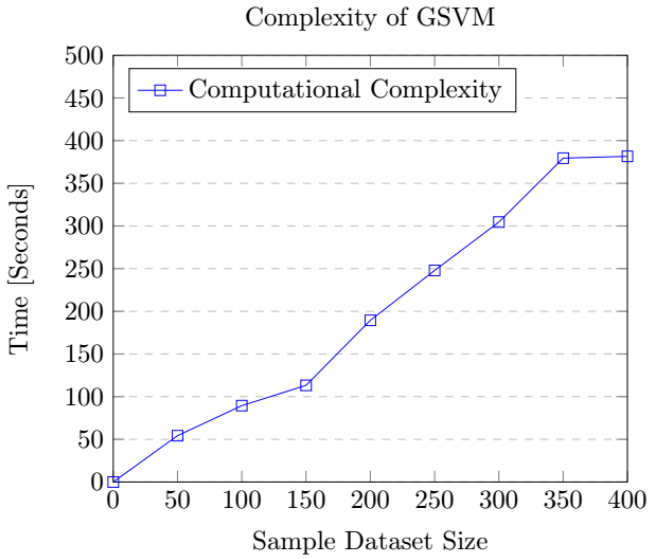

Figure 9: Complexity of the GSVM model

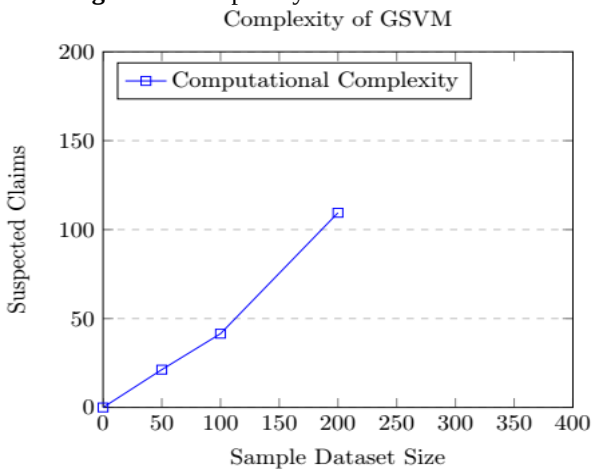

Figure 10: False claims trend

5. CONCLUSION

The aim of this work is to propose a novel model to identify the genuineness of the claims in the agricultural sectors. Normally, the farmer society is the most vulnerable to the natural disasters. Both state and central governments are framing many policies and plans to support farmers of the society. In spite of all this, the farmer suicide rate is gradually increasing. Agricultural insurance is one such step towards the agricultural sectors. But, due to lack of transparency both farmer and insurance companies are cheating to each other. In this paper, by considering the problems of insurance sectors we have proposed a genetic algorithm driven support vector machine-based decision support system to identify the false claims by the farmers. The GSVM segregates the claims based on their genuineness. The proposed model has been simulated over MATAL by using the dataset of KRISHI. During simulation it has been observed that, GSVM able to differentiate the legal and illegal claims with very narrow margin. The recorded performance of GSVM is 81.49% on the considered sample datasets.

REFERENCES

[1] Press Trust of India , "Over 12,000 Maharashtra farmers committed suicide in 3 years: Government," India Today, Maharashtra, 2019.

[2] Prafulla Marpakwar, "150 fewer farmer suicides in 2018," Times of India, MUMBAI, 2019.

[3] India Today, "10 important government schemes in Agriculture sector," India Today Web Desk, New Delhi, 2019.

[4] S. K. Jamanal, K. V. Natikar and S. V. & Halakatti, "Satisfaction Level of Insured Farmers about Crop Insurance Schemes in Northern Karnataka," *Current Journal of Applied Science and Technology*, vol. 38, no. 4, pp. 1-8, 2019. https://doi.org/10.9734/cjast/2019/v38i430370

[5] A. Gulati, P. Terway and S. Hussain, "Crop Insurance in India:Key Issues and Way Forward; Working Paper No. 352," INDIAN COUNCIL FOR RESEARCH ON INTERNATIONAL ECONOMIC RELATIONS, Madhyapradesh, 2018.

[6] Department of Agriculture & Cooperation; Ministry of Agriculture, Government of India., "Report of the Committee to Review the Implementation of Crop Insurance Schemes in India," Krishi Bhawan, New Delhi, 2014.

[7] Vidya, "Viral video claims to expose Bajaj Allianz scam against Maharashtra farmers," India Today, Maharashtra, Mumbai, 2019.

[8] Poorvi Kulkarni, "Why farmers across India are complaining of being cheated by prime minister's crop insurance scheme," Scroll in, Madhya Pradesh, 2019.

[9] Rituraj Tiwari, "Heavy crop-damage payouts likely by insurance firms," ET Bureau, New Delhi, 2019.

[10] Rahul Wadke, "Insurers deploy drones to check claims by farmers," The Hindu, Mumbai, 2018.

[11] H. Mohapatra and A. K. Rath, "Detection and avoidance of water loss through municipality taps in India by using smart taps and ICT," *IET Wireless Sensor Systems*, vol. 9, no. 6, pp. 447 - 457, 2019. https://doi.org/10.1049/iet-wss.2019.0081

[12] L. Di, E. G. Yu, L. Kang, R. Shrestha and B. Yu-qi, "RF-CLASS: A remote-sensing-based flood crop loss assessmentcyber-service system for supporting crop statistics and insurance decision-making," *Journal of Integrative Agriculture*, vol. 16, no. 2, p. 408–423, 2017. https://doi.org/10.1016/S2095-3119(16)61499-5

[13] F. R. Young, O. Chandler and a. A. Apan, "Crop Hail Damage: Insurance Loss Assessment Using Remote Sensing," *Conference Proceeding*, 2004.

[14] P. Reddy and G. Reddy, "Improving the Performance of Loan Recovery and Financial Inclusion for Rural Financial Services with eSaguTM," *Seminar on Rural eEmpowerment Regional Training College, NABARD, Mangalore.*, 2007.

[15] A. Apan, O. Chandler, F. Young1 and T. Maraseni, "Opportunities and limitations of remote sensing for crop loss (hail damage) assessment in the Insurance industry," *Research Gate*, 2005.

[16] R. Navalgund and S. Ray, "Application of space technology in agriculture: an overview," *Smart Agri Post*, vol. 4, no. 6, pp. 6-11, 2019.

[17] M. T. Hasan, "Prospects of Weather Index-Based Crop Insurance in Bangladesh," *International Journal of Agricultural Economics*, vol. 4, no. 1, pp. 32-40, 2019. https://doi.org/10.11648/j.ijae.20190401.15

[18] S. K. Dubey, D. Mandloi, A. S. Gavli, A. Latwal, R. Das and S. S. Ray, "Quality Checking of Crop Cutting Experiments Using Remote Sensing Data: A Case Study For Rice Crop In Odisha," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vols. XLII-3, no. W6, pp. 461-466, 2019.

[19] M. Cerf, M.-H. Jeuffroy, L. Prost and J.-M. Meynard, "Participatory design of agricultural decision support tools: taking account of the use situations," *Agronomy for Sustainable Development*, vol. 32, p. 899–910, 2012.

[20] D. C. Rose et. al, "Decision support tools for agriculture: Towards effective design and delivery," *Agricultural Systems*, vol. 149, pp. 165-174, 2016. https://doi.org/10.1016/j.agsy.2016.09.009

[21] J. Wroblewski, "Theoretical foundations of order-based genetic algorithms," *Fundamental Informaticae*, vol. 28, no. 3-4, pp. 423-430, 1996.

[22] M. Melanie, "An Introduction to Genetic Algorithms," *The MIT Press*, vol. 1, 1998.

[23] F. Lu and J. E. Boritz, "Detecting fraud in health insurance data: learning to model incomplete

Benford's Law distributions," *Machine Learning,* pp. 633-640, 2005.

[24] J. F. Roddick, P. C. J. Li and P. J. Kennedy, "Data mining and analytics 2008," *Proceedings of the 7th Australasian Data Mining Conference*, vol. 87, pp. 105-110, 2008.

[25] Y. Singh and A. S. Chauhan, "Neural networks in data mining," *Journal of Feoretical and Applied Information Technology*, vol. 5, no. 1, pp. 37-42, 2009.

[26] M. Ongayev et. al, "Engineering and Process Infrastructure of the Agro-Industrial Complex," *International Journal of Emerging Trends in Engineering Research,* vol. 7, no. 12, pp. 879-885, 2019.
https://doi.org/10.30534/ijeter/2019/257122019

[27] H. Mohapatra, S. Panda, A. K. Rath, S. Edalatpanah and R. Kumar, "A Tutorial on PowerShell Pipeline and its Loopholes," *International Journal of Emerging Trends in Engineering Research,* vol. 8, no. 4, pp. 975-982, 2020.

[28] N. K. Gattim et. al, "Plant Leaf Disease Detection Using SVM Technique," *International Journal of Emerging Trends in Engineering Research,* vol. 7, no. 5, pp. 634-637, 2019.
https://doi.org/10.30534/ijeter/2019/367112019

[29] N. C. S. Reddy et. al, "An Emperical Study on Support Vector Machines for Intrusion Detection," *International Journal of Emerging Trends in Engineering Research,* vol. 7, no. 5, pp. 383-387, 2019.
https://doi.org/10.30534/ijeter/2019/037102019

[30] H. Mohapatra, "Network Based Intrusion Detection," *Researchgate,* 2017.