# ADVANCED CLUTTER MITIGATION METHOD FOR SURVEILLANCE RADAR USING MACHINE LEARNING

Malwinder Singh, Shashi Ranjan Kumar and Bhukya Soumya Mishra

Product Development & Innovation Center, Bharat Electronics Limited,
Bengaluru, India

### *ABSTRACT*

*This research focuses on improving the ground clutter mitigation by integrating ML methods with traditional methods (such as CFAR and Doppler processing) of X-band surveillance radar. Discriminative machine learning methods are used as they have the ability to learn without the knowledge of distribution type. The techniques used to accomplish research includes raw IQ radar data collection, data labelling, and feature generation, statistical significance of generated features, model (DT, SVM and ANN) training and model evaluation. The results indicate improvement in mitigation of ground clutter for different scenarios. The research also discusses the future work related to this research.*

### *KEYWORDS*

*Artificial Neural Network (ANN), Constant False Alarm Rate (CFAR), Digital Down Conversion(DDC), Decision Tree (DT), Fast Fourier Transform (FFT), false negatives (FN), false positives (FP), In Phase-Quadrature Phase (IQ), Machine leaning (ML), Support Vector Machine (SVM), true negatives (TN), true positives(TP)*

## 1. INTRODUCTION

The traditional signal processor consists of CFAR and Doppler based signal processing, which relies on two key features for detection decision making, the detection-threshold and Doppler. It applies Fast Fourier Transform (FFT) and uses Doppler information to separate target from clutter as well as noise, in frequency domain. Further, for each frequency bin, CA-CFAR produces a detection threshold based on power. However, detection based on this approach have limited accuracy in heterogeneous ground clutter environment. In particular, CA-CFAR assumption of Gaussian distribution [3] for noise power is no longer valid due to presence of clutter. Instead, ground clutter for X-band surveillance radar follows the Weibull distribution and log-normal distribution [2]. The discriminative machine learning methods can learn without the knowledge of distribution type. Therefore, machine leaning methods such as DT, SVM and ANN can be used to discriminate between target and clutter. This research aims to extract more features from raw IQ radar data, test hypothesis about these features, and application of these features to machine learning models to attain the objective of improved clutter mitigation.

### 1.1. Motivations

The existing process is based upon CFAR, which is equivalent to generative model based on Gaussian distribution. The discriminative models are more accurate compared to generative

David C. Wyld et al. (Eds): ICAITA, CDKP, CBW, SOFT, CMC, Signal, MLT-2023
pp. 117-127, 2023. CS & IT - CSCP 2023 DOI: 10.5121/csit.2023.130609

models. Another motivation is that the existing process uses only two features (i.e. CFAR threshold and Doppler). It is more likely to linearly separate data in higher dimension. The research is aimed to create more features where target can be separated more accurately from clutter.

### 1.2. Objectives

The objective of present study is to improve target detection by using machine learning as an alternate to conventional detection process for X-band perimeter surveillance radar.
Specific objectives are:

- Feature extraction from raw IQ radar data.
- Hypothesis testing for extracted features.
- Train different machine learning models.
- Selecting most optimal model for integration to existing method.

### 1.3. Main Contributions

The main contribution of this research is to implement a unique ML approach for radar target detection for X-band perimeter surveillance radar, which can distinguish between target and ground clutter better than the conventional signal processing method and have prediction time within a single Coherent Pulse Interval. This new approach uses different features extracted from raw IQ radar data.

## 2. Data and Methodology

For better accuracy of models, the real time radar IQ data is collected for multiple scans with walking man as the controlled target and then the required processing is done. To remove the bias, data collection is performed over different sites, different hardware and different time. The collected data is stored as PCAP file on a local host, which is further used for extracting raw data bytes and stored as raw file. The raw is imported in MATLAB and radar data cubes are extracted, which further used for new feature generation [1]. Next step is the data labelling for supervised learning. The labelled features are analysed through correlation and P-value hypothesis testing. Based upon analysis results, features are selected for model training. The selected features dataset is cleaned for missing values and used for training of 3 binary classification machine learning methods, which are DT, SVM and ANN. All the models are evaluated through a technique similar to k-fold validation. Finally, comparison among the models is performed based on accuracy and confusion matrix

### 2.1. Data Collection & Preparation

A walking man is used as the controlled target and radar received IQ data is collected. The radar signal processor unit and other sub-systems are connected by the Ethernet switch. The signal processor converts radar echoes to IQ pair and sends to local host via UDP protocol, where these packets are stored as PCAP file. Data collection is performed using X-band surveillance radar at 2 different locations and Google Map view is shown in fig1. The yellow lines mark clutter boundary, and red lines mark coverage. The equivalent clutter seen by radar for each site on decibel power scale is shown in fig 2.

Fig. 1. Google Map View of Data Collection Sites

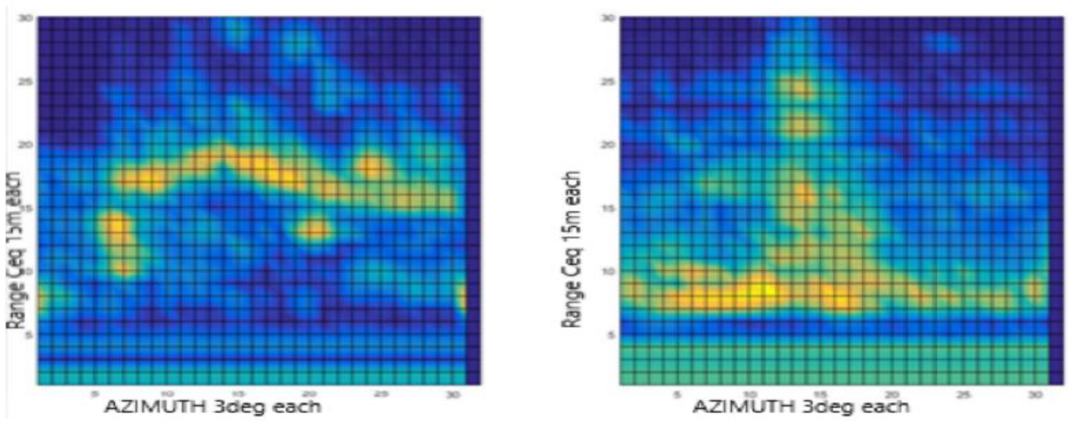


Fig. 2. Radar map view of the clutter image (dB scale)

The controlled target is used in 2 type of strategies as shown in the fig 3. This is to cover radial and tangential target motion conditions, and also to make data labelling easy. To remove the bias in the data collection process, the data is collected with different hardware, different physical locations and different time (days) for total of 438 radar scans.

WIRESHARK Ethernet capturing tool is utilized for capturing data packets and saving in a file. Each packet represents a single pulse data and contains 100 range cells IQ data for 2 channels. The WIRESHARK tool is used to convert the PCAP file into RAW file, which can be processed directly by MATLAB.

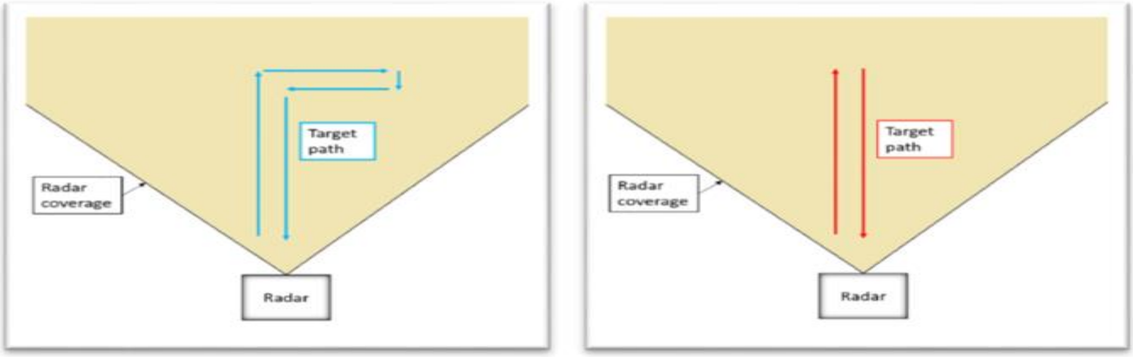


Fig. 3. Controlled Target Path Strategy

## 2.2. Feature Generation

The main notion behind feature generation is to create features for which clutter and target class has statistically different value. The MATLAB tool is used for feature generation [1]. The "DETECTION" feature is considered to contain both clutter and target. The traditional signal processing is used to generate DETECTION (represented by +1 and -1 in final dataset). Some of the feature are generated during pulse integration across the slow time axis, while the remaining features are generated by integrating over the scans. The brief idea about generated features is given below:

– FDM THRESHOLD: In this feature, for each Doppler band, moving average power over the scans is calculated.
– FAN THRESHOLD: The FAN threshold represents the probability of a range cell being clutter or target. This exploits the static nature of clutter in space over the scans.
– POWER DECAY(PD): The power ratio between range-cell to next range-cell. i.e. abs(Xn) / abs (Xn + 1)
– SIGMA: This feature represents Doppler spectrum width of a range cell [8].

$$Sigma = \frac{\lambda}{2 * \sqrt{2} * \mathit{\Pi} * T} * \sqrt{ln\left|\frac{R(0)}{R(1)}\right|}$$

Where T is PRT of Radar, R(0) & R(1) is first & second moment respectively and λ is wavelength of radar operating frequency.

– SNR: The signal to noise ratio represents the peak power of signal to average noise floor ratio, after removal of zero Doppler bins from FFT.
– DYS: This is complex ratio of DIFF channel to SUM channel, used as indication of angular error in mono pulse application [10].
– DETECTION: This detection from the traditional method (CA-CFAR and FFT). It contains true target as well as clutter. Indicate as +1 or -1;

The generated features are saved as 2D matrix in MATLAB for each scan. A sample of features for a scan is shown in fig 4.

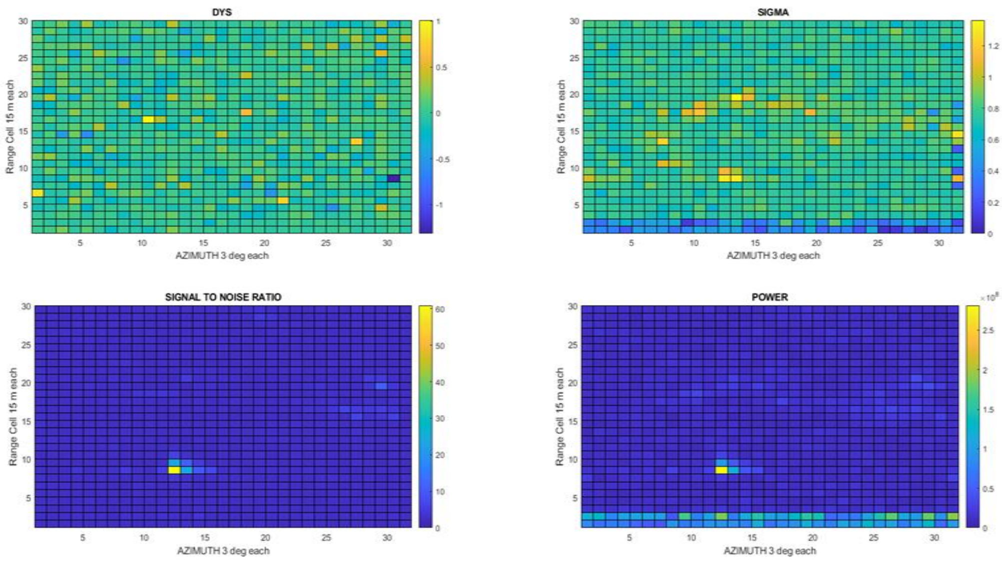

Fig. 4. Features 2D matrix view of a scan

## 2.3. Data Labelling

The supervised learning requires labelled dataset. The manual labelling is done after the generation of all features is completed. The labelled data is stored as 2D matrix for each scan similar to features matrix. The following factors are considered for labelling a particular range-cell as a TRUE LABEL:

– Detection range-cell from conventional method (CA-CFAR, FFT)
– Alignment to expected target path.
– Location of Fixed clutter area. (From clutter map).
– Previous location of LABEL (target) range-cell.
– Intuition from other features.

Based upon above factors, the TRUE LABEL is decided for every range cell. For valid target, LABEL=1, otherwise LABEL=0. This process produces a 2D MAT file (as shown in fig 5) of 30 range cells x 32 Azimuth. The MAT files are 2D files, and easy to process in MATLAB. However, in python programming, CSV file can be handled more efficiently. Therefore, all the 2D data from MAT file is converted into 1D data, and written to CSV file azimuth by azimuth. The features represent columns of the CSV file. This process is repeated for each datasets and CSV files are generated per dataset.

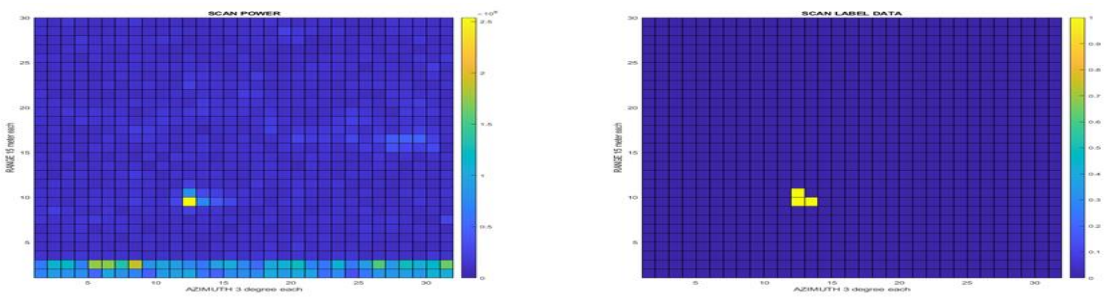

Fig. 5. Labelled 2D matrix view

## 2.4. Feature Analyses

The all datasets are converted in CSV format with labelled data. These datasets are imported in python and combined into single data frame. The data contains NaN values for some features. This is due the reason of partial 1st scan or feature property of being calculated across the scans, where one has to wait for a whole scan to generate the value. The NaN values are simply dropped, as radar is a real time system and value imputation is not applicable. One of the main issues with Radar Signal Processor dataset is that there is huge imbalance between the classes. Which is due to practical reason that, out of total range-cell (30 X 32) of a scan, only small percentage represents the controlled target. The plot shows the imbalance of classes in fig 6.

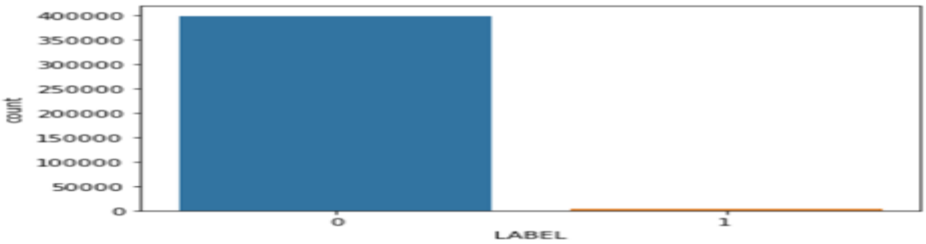


Fig. 6. class imbalance plot

The data frame contains the following range-cell types:

– Non detections: DETECTION = -1 & LABEL=0.
– Clutters: DETECTION = 1 & LABEL=0.
– Targets: LABEL=1.

To get around this problem, the data frame is filtered (as shown in fig 7). The average count of non-detection, clutter and targets is calculated and it was found statistically that proportion of each type is [92.27% 7.01% 0.71%] respectively. The LABEL data class (true target) less than 1%. Major data is non-detection class type, which comes from noise range cells. This data can be filtered by existing method (CA-CFAR) as well, but the clutter properties are very similar to that of target and cannot be mitigated in the existing method. Therefore, the non-detections are removed from data frame and only clutter and target data are retained. As the main purpose of the research is to improve the target detection in the presence of clutter.

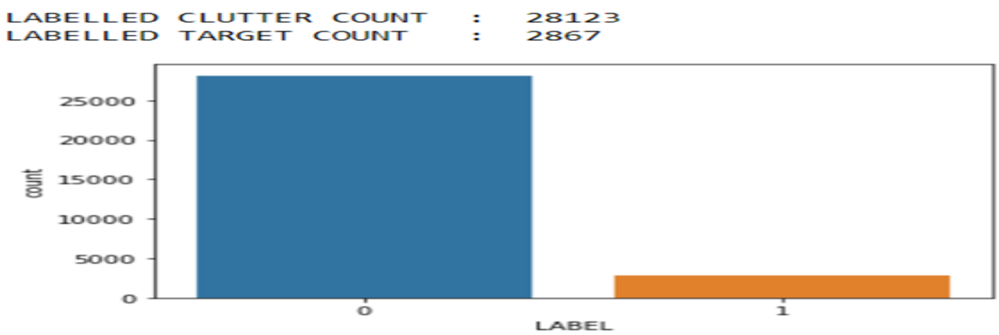


Fig. 7. Class imbalance after filtering non-detections

The data corresponding to each class is described to understand the central tendency and compare the two classes. The Fig 8 shows the target class description and Fig 9 represents the properties of clutter class. It is evident that each class central tendencies for new generated features are different.

| | LABEL | DETECTION | FDM_TH | FAN_TH | PD | DYS | SIGMA | SNR |
|---|---|---|---|---|---|---|---|---|
| count | 2867 | 2867 | 2867 | 2867 | 2867 | 2867 | 2867 | 2867 |
| mean | 1 | 0.365888 | 2.002259 | -7.126962 | 3.616943 | 0.07325 | 1.057315 | 47.03747 |
| std | 0 | 0.930821 | 1.505701 | 8.900155 | 2.841572 | 0.411826 | 0.386796 | 44.3873 |
| min | 1 | -1 | 0.0797 | -15 | 0.053632 | -1.6174 | 0.003133 | 1.7865 |
| 25% | 1 | -1 | 0.81839 | -14 | 1.53925 | -0.13291 | 0.806875 | 13.6245 |
| 50% | 1 | 1 | 1.6345 | -10 | 2.844 | 0.032823 | 1.0246 | 35.94 |
| 75% | 1 | 1 | 2.83375 | -4 | 5.11105 | 0.23674 | 1.32125 | 65.542 |
| max | 1 | 1 | 8.4424 | 15 | 17.509 | 3.1814 | 2.7267 | 366.34 |

Fig. 8. target class central tendencies

The further analysis done for redundancy elimination. The features redundancy is tested using correlation test among the features. The correlation results are shown in fig10. The test is performed by further balancing the class through random sampling method.

From the correlation results:

– No two feature has correlation outside ±0.4 range.
– All features can be selected

| | LABEL | DETECTION | FDM_TH | FAN_TH | PD | DYS | SIGMA | SNR |
|---|---|---|---|---|---|---|---|---|
| count | 28123 | 28123 | 28123 | 28123 | 28123 | 28123 | 28123 | 28123 |
| mean | 0 | 1 | 2.090699 | 9.365893 | 4.383549 | 0.004643 | 0.802069 | 4.866778 |
| std | 0 | 0 | 0.986869 | 9.311752 | 3.44872 | 0.210406 | 0.32986 | 7.761027 |
| min | 0 | 1 | 1 | -14 | 0.024749 | -3.2447 | 0.008888 | 1.2061 |
| 25% | 0 | 1 | 1.3247 | 6 | 1.19135 | -0.08699 | 0.63826 | 2.4242 |
| 50% | 0 | 1 | 1.8331 | 15 | 3.808 | -0.00324 | 0.76845 | 2.6189 |
| 75% | 0 | 1 | 2.57475 | 15 | 6.94045 | 0.080475 | 0.94976 | 3.0464 |
| max | 0 | 1 | 7.4419 | 15 | 22.349 | 2.8733 | 2.6784 | 164.45 |

Fig. 9. clutter class central tendencies

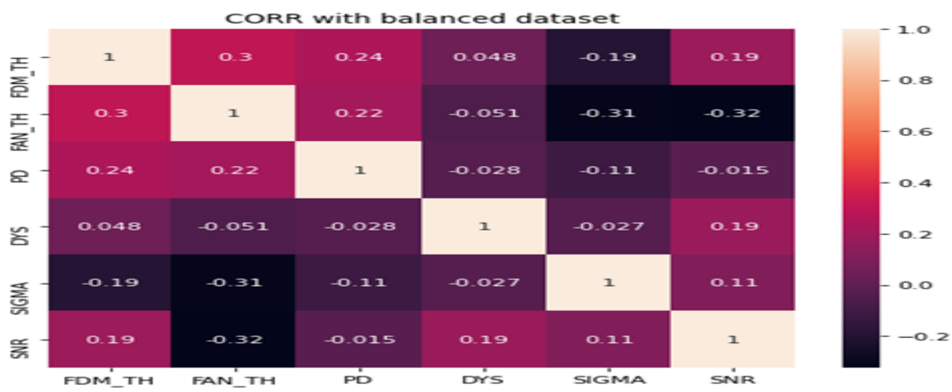


Fig. 10. correlation between features

Final analysis is performed for significance test and results are shown in Fig 11. The P-value test is conducted for all the features to find out, whether difference in the μ value of two classes is, of any statistical significance. The P-value is observed 12.5% for SNR feature, and for remaining features, the P-value is less than 2%, which is quite significant.

| Feature | P-value | Z-value |
|---|---|---|
| FDM_TH | 1.91E-02 | -2.072552 |
| FAN_TH | 1.23E-14 | -7.623709 |
| PD | 1.11E-03 | -3.058678 |
| DYS | 6.59E-96 | 20.746496 |
| SIGMA | 0.00E+00 | 75.697125 |
| SNR | 1.25E-01 | 1.152524 |

Fig. 11. P-value statistics results

### 2.5. Model Training & Evaluation

Model development is carried out in python using Google Colab tool and all the models are imported from sklearn machine learning library. The available data for training and evaluation, contains 28123 clutter class data samples and 2867 target class samples. The class imbalance problem is taken care by using a custom technique very similar to k-fold validation. The clutter class is divided into 9 sub-groups where each sub-group size is equal to target class group. The 9 different experiments are created. For each experiment, using random selection, the training and test data split with proportion of 70:30 respectively. The model is trained on training samples (70%) and evaluated on test samples (30%). The process is repeated 9 times for each model type.

In case of DT model[12], gini index is used as the entropy measure. To avoid the over-fitting, maximum tree depth constraint to 5 and maximum leaf node constraint to 8 nodes. For the SVM [9] model, linear kernel is used to separate the clutter class from target labelled class, as linear kernel requires less training time. In case of the ANN model [11], the shallow neural network is used which has 2 hidden layers with each layer having depth of 4 neurons. The relu activation function used in the hidden layers, whereas sigmoid is used in the output layer. Each model type is trained and evaluated for total 9 experiments. The accuracy is calculated as minimum, average and maximum accuracy for each model.

## 3. Results and Discussion

The validation results of all 3 models are given in Fig 12. The accuracy for each model is averaged over multiple experiments to avoid the bias condition. Comparison table shows that decision tree (DT) has the least average accuracy among the 3 models, and the ANN model has the highest average accuracy.

| Model | Minimum accuracy | Average accuracy | Maximum accuracy |
|---|---|---|---|
| DT | 0.8082 | 0.8762 | 0.9680 |
| SVM | 0.8332 | 0.9076 | 0.9709 |
| ANN | 0.8448 | 0.9138 | 0.9715 |

Fig. 12. model accuracy matrix

The confusion matrix is aggregated over multiple experiment for all the models and results are shown in Fig 13. The important aspect of confusion matrix in our use case is FN value for each model, which indicate the true target drop ratio. This ratio is about 6% for SVM and ANN, but more in case of DT and is about 11%. This criterion is important for model selection.

| | Confusion matrix | |
|---|---|---|
| DT | 6761(TN) | 1033(FP) |
| | 883(FN) | 6812(TP) |
| | | |
| SVM | 6834(TN) | 960(FP) |
| | 471(FN) | 7224(TP) |
| | | |
| ANN | 6970(TN) | 824(FP) |
| | 511(FN) | 7184(TP) |

Fig. 13. confusion matrix results

The main concept of training multiple models is to select the most optimal model. The purpose of this research is not just to create model using python library, but also to integrate the equivalent model with existing signal processing technique. For this reason, DT is selected as most optimal model, considering the accuracy results and efforts required to integrate with existing signal processing unit running on FPGA based system. The DT model has the average accuracy (87%) close to other models (90% & 91%) and can be directly written in VHDL with nested if-else structure.

## 4. LIMITATIONS OF THE RESEARCH

The models developed by current research have limitations when target is moving tangential near clutter boundary where performance of the models may be compromised. The fig 14 shows the average accuracy drop for each model when tested with tangential motion data. This data filtered from the main data where target motion is tangential near the clutter boundary. From the results, the performance impact is least in case of SVM, and most in case of DT.

| Test scenario | DT | SVM | ANN |
|---|---|---|---|
| **Regular test data** | 0.8468 | 0.8759 | 0.8834 |
| **Tangential data** | 0.7556 | 0.8036 | 0.7943 |
| **Accuracy drop** | 0.0912 | 0.0723 | 0.0891 |

Fig. 14. Model limitations

## 5. CONCLUSIONS

The correlation results and the P-value results for the generated features are statistically significant and can be used to discriminate between clutter and target detection. The clutter mitigation for all three models (DT, SVM and ANN) have average accuracy better than 87%. The clutter mitigation accuracy with traditional method is observed less than 10%. Hence, it is very much evident that proposed machine learning approach, integrated with traditional signal processing techniques, has significant improvement in clutter mitigation for X-band perimeter surveillance radar ground clutter scenarios.

## 6. FUTURE WORK

The one aspect for future work is to use MATLAB as modelling tool. The model developed in this research can be regenerated in MATLAB and converted into VHDL synthesizable model using HDL Coder, which can run on the real time FPGA based embedded hardware of signal processor for better prediction-speed and integrity. Also the data collection method can be improved by embedding the feature generation portion inside embedded hardware instead of collection of raw IQs. The raw IQ collection requires huge data rate and memory. Reports are generated only for detections, which reduces the size of data by huge margin. Finally, data labelling aspect of this research is manual and institution based. The labelling technique can be explored for automation.

## ACKNOWLEDGEMENTS

We are thankful to all the co-workers, reporting officers of all the team members at BEL, Bangalore who have given their valuable inputs for this study. The data collection facility provided by BEL and the guidance received from senior officer–Ms. Sarala (AGM, PDIC/Radar System)–was invaluable, the team is highly thankful to all of them and also to the top management for providing continuous encouragement and opportunities to express our views.

## REFERENCES

[1] Radar systems analysis and design using matlab® third edition Bassem R. Mahafza deciBel Research Inc. Huntsville, Alabama, USA

[2] Statistical Analyses of Measured Radar Ground Clutter Data J. B. Billingsley, Member, IEEE MIT Lincoln Laboratory A. Farina, Senior Member, IEEE Alenia Marconi Systems F. Gini, Member, IEEE M. V. Greco L. Verrazzani University of Pisa April 1999

[3] Y. Wang, W. Xia, and Z. He, "CFAR knowledge-aided radar detection with heterogeneous samples,"IEEE Signal Process. Lett., vol. 24, no. 5, pp. 693697, May 2017.

[4] Research on a New Comprehensive CFAR (Comp-CFAR) Processing Method YI LIU, (Fellow, IEEE),Shufang Zhang, Jidong Suo, Jingbo Zhang, And Tingting Yao Institute of Information Science and Technology, Dalian Maritime University, Dalian 116026, China, Feb-2019

[5] Robust CFAR Detection for Multiple Targets in K-Distributed Sea Clutter Based on Machine Learning Jiafei Zhao, Rongkun Jiang *, Xuetian Wang and Hongmin Gao School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China; zhaojiafei@bit.edu.cn (J.Z.); wangxuetian@bit.edu.cn (X.W.); gaohm@bit.edu.cn (H.G.) May-2019

[6] A Comprehensive Survey of Machine Learning Applied to Radar Signal Processing Ping Lang, Xiongjun Fu, Marco Martorella, Jian Dong, Rui Qin, Xianpeng Meng, Min Xie Sep-2020

[7] Target Detection using Radar Processors based on Machine Learning Enrique V. Carrera; Fernando Lara; Marcelo Ortiz; Alexis Tinoco; Rub´en Le´on 2020 IEEE Andescon Dec-2020

[8] Bringi, V. N., and V. Chandrasekar, 2004: Polarimetric Doppler weather radar: Principles and applications. Cambridge University Press, pp 282.

[9] A Comparison of SVM Kernel Functions for Breast Cancer Detection, Muhammad Hussain, Summrina Kanwal Wajid, 2011 Eighth International Conference Computer Graphics, Imaging and Visualization

[10] Monopulse Radar Target Detection in the Case of Main-Lobe Cover Jamming by Lei Wang, Qiliang Zhang and Pengfei Wan

[11] ADAM: A M ETHOD FOR S TOCHASTIC OPTIMIZATION Diederik P. Kingma University of Amsterdam, Open AI Jimmy Lei Ba University of Toronto

[12] Introduction to Machine Learning, Neural Networks, and Deep Learning Rene Y. Choi, Aaron S. Coyner, Jayashree Kalpathy-Cramer, Michael F. Chiang, and J. Peter Campbell

## AUTHORS

**Malwinder Singh (Manager, BEL)**

Completed his B.E from Thapar University, Patiala, Punjab. He joined BEL in December 2010 & currently working as Manager in Radar Signal Processing Department. He also completed his M.Tech from BITS-Pilani in Data science. He has worked in design and development of signal processor of Pulsed radars. His current area of work is design of FPGA based PCB design, MATLAB, test GUI development and RTL based software development for radar signal processors.

**Shashi Ranjan Kumar (Manager, BEL)**

Completed his B.E from Muzaffarpur Institute of Technology under B.R.A. University, Bihar. He joined BEL in September 2007 & currently working as Manager in Radar Signal Processing Department. He also completed his M.Tech from DIAT, Pune in signal processing& communication & was awarded Gold medal. He is involved in design & development of digital receiver & signal processor for pulse & CW Radars. His current area of work is design of FPGA based compact & power efficient signal processor for Radars.

**Bhukya Soumya Mishra (Deputy Engineer, BEL)**

Completed her B.Tech from National Institute of Technology Calicut, Kerala. She joined BEL in February 2021 & currently working as Deputy Engineer in CoE - (Radar & Weapon System) Department. Her current area of work is MATLAB, LABVIEW, FPGA based PCB design and RTL based software development.