Deep learning-based explainable target classification for synthetic aperture radar images

Mandeep, Husanbir Singh Pannu
Computer Science and Engineering Department
Thapar Institute of Engg. & Tech.
Patiala, India
mmandeep_be19@thapar.edu, hspannu@thapar.edu

Avleen Malhi
Computer Science Department
Aalto University
Finland
avleen.malhi@aalto.fi

***Abstract*—Deep learning has been extensively useful for its ability to mimic the human brain to make decisions. It is able to extract features automatically and train the model for classification and regression problems involved with complex images databases. This paper presents the image classification using Convolutional Neural Network (CNN) for target recognition using Synthetic-aperture Radar (SAR) database along with Explainable Artificial Intelligence (XAI) to justify the obtained results. In this work, we experimented with various CNN architectures on the MSTAR dataset, which is a special type of SAR images. Accuracy of target classification is almost 98.78% for the underlying pre-processed MSTAR database with given parameter options in CNN. XAI has been incorporated to explain the justification of test images by marking the decision boundary to reason the region of interest. Thus XAI based image classification is a robust prototype for automatic and transparent learning system while reducing the semantic gap between soft-computing and humans way of perception.**

***Index Terms*—Artificial intelligence; deep learning; image classification; target recognition; synthetic aperture radar**

I. Introduction

Artificial Intelligence uses deep, distributed computational architectures to solve the real world complex problems. Real dataset often suffers from noise and artifacts, so the recognition process carried out with the help of abstract level learning methodologies. A lot of these initiatives are biologically inspired due to the fact that the human brain acquires most of its practical and logical reasoning capability by processing in this way. Hence, latest advances in algorithms and computation have focused attention on a new class of biologically inspired algorithms introduced as Deep Neural Networks (DNNs) [1]. There are an enormous number of layers with thousands of nodes interconnected with each other analogous to the brain with an extensive network of neurons. The major application of these networks lies in classification decisions with the main advantage of its learning capability of complicated decision functions compared to other techniques. Again, these models must be able to give justification about the model rationale which can be evaluated by experts to audit the decision making factors. There should be a measure to see how the machine reasons for an outcome in contrast to a human expert for potential conflicts and legal norms.

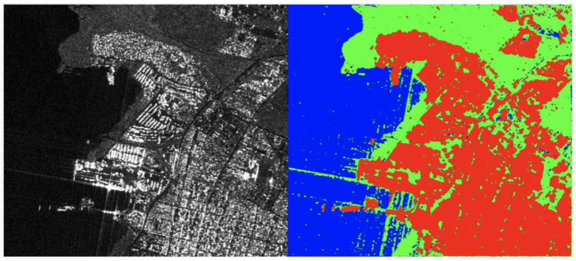

Fig. 1. Left: Initial SAR image of Port-au-Prince (Haiti) (©ISA, 2009). Right: Classification map obtained with the hierarchical method for the 3 classes (Blue: water; Green: vegetation; Red: urban area) [2]

The definition of explainability of artificial intelligence is that it is a formal explanation by a model against action taken or decision made, given the test data and features involved.

A. Target recognition using SAR

Synthetic Aperture Radar (SAR) in Automatic Target Recognition (ATR) problems is a common application of such networks. The ability of constant surveillance provided by SAR has made it an irreplaceable imaging radar technology. SAR can provide images of land, sea and air targets during all weather conditions. This is comparable to the problem of image classification with a huge labeled images' database to fulfill the prerequisite of labelling a new anonymous image. A distinct amount of characteristics actually distinguish SAR modality from natural imagery, most importantly the fact that both magnitude and phase are included in the data. SAR is particularly useful for tasks such as remote sensing, surveillance, reconnaissance and target recognition. Analysts are trained to understand and exploit the raw SAR data for the identification of targets of interest and significant activities. Limitations in radar technologies restrict image resolutions to tens of centimeters or meters per pixel. Thus, exploiting the SAR images becomes a complex process and requires years of training of image analysts. This is because they manually search and classify targets that extend for just a couple of meters in large SAR images that covers tens of kilometers. The time requirement of this manual classification is significant and reduces the performance of the intelligence agencies. They generate massive amounts of data and demand

TABLE I
COMPARATIVE ANALYSIS OF THE STATE-OF-ART ON MOVING AND STATIONARY TARGET ACQUISITION AND RECOGNITION (MSTAR) DATASET

Sr.	Reference	Year	Technique	Accuracy
1	Coman [3]	2018	CNN; Layers: 2 Conv, 1 flattened, 2 dense, 2 dropout	90%
2	H. Furukawa [4]	2017	CNN; Layers: 17 Conv, 1 FC, Based on ResNet-18 [5]	99.56%
3	S. Zaied [6]	2018	Architecture 1:CNN Layers: 2 Conv, 1 hidden Architecture 2:CNN + CAE Layers: 2 Conv, 1 hidden	Arch. 1: 75.98% Arch. 2: 90.09%
4	Z. Lin [7]	2017	Ensemble; Models: 2 CHU	99.09%
5	I. M. Gorovyi [8]	2017	SVM	90.07%
6	R. Min [9]	2019	MCNN; Layers: 1 Conv, 1 FC. Teacher Network: ResNet-18 [5]	98.2%
7	R. Chakraborty [10]	2019	CRN; Layers: 2 C Conv,2 G-transport, C Residual, Invariant Layer, 3 Conv, 3 Batch Norm and ReLU, 2 R R Block, 1 MaxPool, 2 FC	97.69%
8	Q. Liu [11]	2018	ConNet; Layers: 4 Conv, 1 Conv Filter, 1 FC	99.48%
9	M. Heiligers [12]	2018	CNN: Layers: 4 Conv + ReLu, 2 MaxPool, 1 FC, 1 Soft-max	97.6%
10	**Proposed**	**2020**	**CNN: Layers: 3 Conv + ReLu + Batch Norm, 2MaxPool, 1 flattened, 1 Dense + Softmax**	**98.78** %

the customized algorithm which is easy to implement and well generalized. Considering the facts, deep learning algorithms is an ideal fit for automatic feature extraction and target classification. Consequently, the need for ATR algorithms for radar images has made it an active research area for many years. The research community of deep learnign domain have adopted SAR ATR as one of the benchmark problems for highlighting the potential of these new methods. The Moving and Stationary Target Acquisition and Recognition (MSTAR) database [13] is a publicly available dataset formed by a collection of eight military vehicles taken from a number of aspect angles which can be employed broadly for algorithm development and consistent performance comparison. This paper is organized as follow: section I is introduction; section II is about survey of recent literature for research motivation; section III is proposed method; section IV is experimental analysis and section V is conclusion.

II. LITERATURE REVIEW

CNN is used for the classification of the Synthetic Aperture Radar images on datasets like MSTAR in [3]. The classification is done with and without additional radar information. The results are then compared with the performance of traditional ML models. In [4], CNN has been used to classify SAR imagery with and without data augmentation along with translation invariance of CNN. The accuracy has been found to be 99.6% on MSTAR dataset with 10 classes. Translation invariance has been introduced in the MSTAR dataset as a form of data augmentation techniques and with the help of Accuracy-translation map and plots. Further, CNN and Convolutional autoencoders are used to classify SAR and Inverse SAR images from the MSTAR dataset of ten classes such that the CAE provides optimal filters to the CNN layers for the classification of the dataset [6]. The problem of limited availability of publically available SAR imagery in the MSTAR dataset is tackled by the use of the Convolutional highway unit and the use of an ensemble model that consists of two CHU-Net to generate multiscale feature representations of SAR images [7]. The use of the SVM classifier done on imagery dataset available publically as MSTAR dataset [8]. The use of well-crafted features and proper preprocessing of the image dataset is proposed over the use of CNN as a method to prevent the overfitting of data. In [9], due to the requirement of high memory and bandwidth connection in deployment of Deep CNN in real-time recognition systems of SAR sensors, a micro CNN trained through a deep CNN, is proposed which has the memory footprint that has been compressed 177 times, and the calculated amount reduced by a factor of 12.8. The use of Complex-valued deep learning is proposed for the classification of MSTAR dataset through DNNs defined on the space of complex numbers that utilize weighted Fréchet mean [10]. Compared to its state-of-the-are counterpart on the same dataset the proposed model is able to achieve better performance with the use of just 1% of the parameters. In [11], CNN has been used to construct well-defined features form a limited MSTAR dataset, which is then used as features for the SVM model for the classification of a complete MSTAR dataset. This method of feature extraction from CNN is claimed to be more effective as compared to traditional hand-crafted features for this dataset. The classification of the MSTAR dataset is achieved through the use of CNN [12]. Along with the classification, the decisions of the CNN are explained through the visualization of a saliency map which has been computed with the help of the Grad-CAM technique. The XAI tool, LIME has already been used in many applications to provide the explanation of the black box model decisions for images and textual data [14] [15]. The aim is use the LIME to explain the classification results for SAR images as well. Table I provides the state-of-the-art comparison of the existing techniques on MSTAR dataset.

III. PROPOSED METHOD

The proposed technique is based on CNN architecture as shown in figure 2. The first phase of methodology comprises the MSTAR image dataset collection and CNN is applied for the image classification on SAR images. The detailed architecture for the CNN model has been influenced from [16] and [17] with empirical modifications for best possible validation. The architecture diagram of proposed CNN is depicted in figure 3. In the second phase, an explainable artificial intelligence tool, Local Interpretable Model-agnostic Explanations (LIME) is used to provide the explanations of the image classification

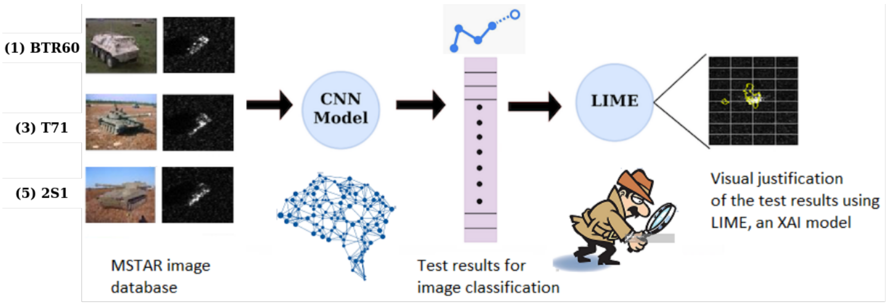
(1) BTR60
(3) T71
(5) 2S1
CNN Model
LIME
MSTAR image database
Test results for image classification
Visual justification of the test results using LIME, an XAI model

Fig. 2. XAI incorporated to CNN predictions for test results justification

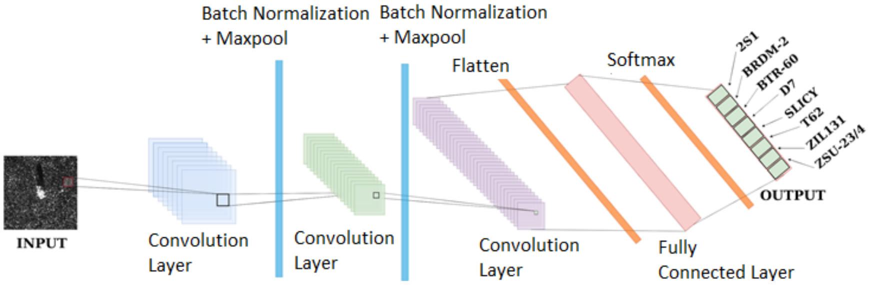

Fig. 3. Proposed CNN architecture for target recongInition

results. LIME [18], [14] is the original Python implementation of one of the explanations techniques used in literature. The neural network generated by TensorFlow acts as input to LIME and results in the matrix representation of the regions triggering the particular classification in the form of a specific frame. LIME enables post-hoc explainability which helps in providing local explanations for a particular decision made by machine learning so that it can be made interpretable on demand rather than explaining the whole systems behavior. The proposed explainable deep learning based image classification is prototype system for automatic and transparent learning system.

IV. Experimental Results

This sections has dataset description, augmentation procedure, performance metrics, XAI for the justification of agnostic CNN model.

A. Dataset description

CNN training has been done using MSTAR dataset which contains 8 classes. It is compiled and processed by the Sandia National Lab and is publicly available[1]. The specifications of the images of each class are described in the Table II.

B. Data Augmentation

The process of data augmentation[2] used in this paper is very efficient and easy to follow. All the images are subjected to a function such that the output image has an equal 0.33 probability of being flipped sideways, inverted, and of no change. The original distribution of the images of different classes is skewed. Some classes contain only 195 images while some have more than 1,900 images. This skewness causes the CNN model to overfit to some classes while underfit for others. To solve this problem the proposed methods has used *two cases*:

TABLE II
Classes count before and after image augmentation

Classes	Before Augment	After Augment
2S1	577	577
BRDM-2	697	697
BTR-60	195	**585**
D7	274	**548**
SLICY	1953	1953
T62	273	**546**
ZIL131	274	**548**
ZSU-23_4	696	696

- Use the same number of images for each class.
- Perform data augmentation for under sampled classes as shown in the Table II.

The *first case* based method reduces the data available for training, validation, and testing considerably with just 170 images per class for training, 15 images for validation and 10 images for testing. When the CNN model was trained on this reduced dataset, the performance was below 80 percent and the model still experienced over-fitting. One of the major reason for this was the fact that during data preprocessing all the images were resized to $200 \times 200 \times 1$ and $200 \times 200 \times 3$. The size of some classes was as low as $54 \times 54 \times 3$ while some classes had images with a size of $198 \times 198 \times 3$. When all the classes we resized, the classes with a low resolution

[1] [Online].Available: https://www.sdms.afrl.af.mil/index.php?collection=mstar

[2] [Online]. Available: https://github.com/aleju/imgaug

before the preprocessing had less information per image before and after the preprocessing. These classes incidentally also have a large number of total images. When we dropped the images from these classes to even out the data, we excluded information related to these images. The effect of this reduction of information had severe effect on the classes with less information per image. The *second case* involved the use data augmentation to even out the dataset distribution. Data augmentation was done to classes with than 250 images. After the data augmentation, CNN model was trained on the complete dataset. This time the model did not overfit the data as the number of training, validation, and test images increased. The reason for specific augmentation just for few clases in Table II is that the dataset for this classification problem is unique as compared to the broad public datasets like ImageNet, standard data pre-processing techniques like data augmentations provide only limited support in improving the performance of the model and overuse of such methods results in over-fitting.

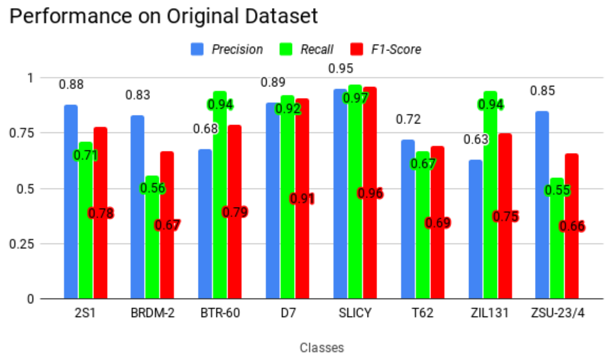

Fig. 4. Performance of CNN on original (unaugmented) dataset

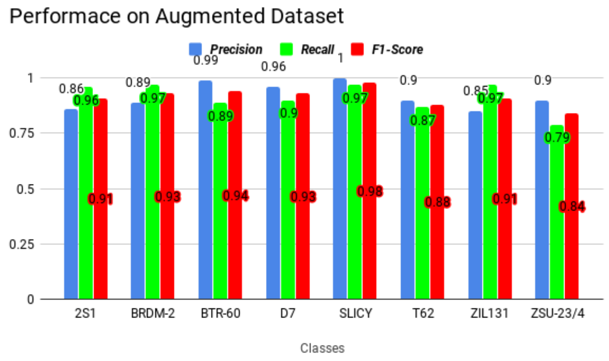

Fig. 5. Performance of CNN on augmented dataset

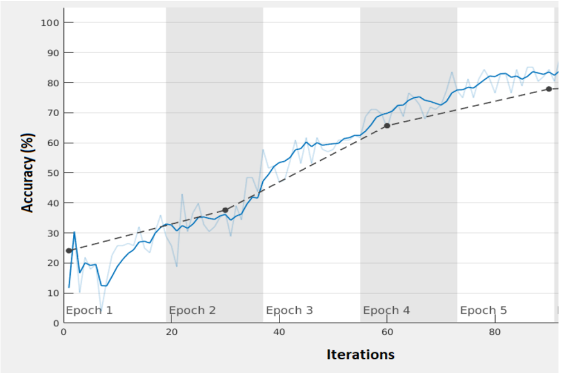

Fig. 6. Accuracy 80%

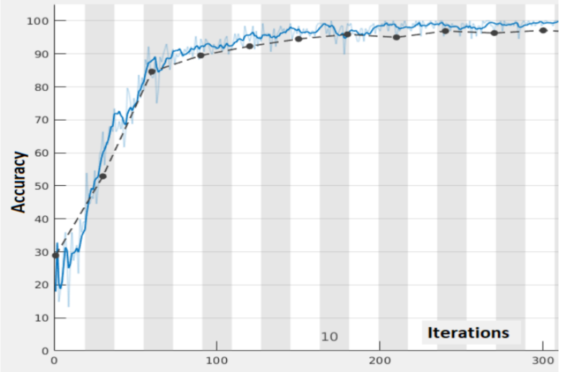

Fig. 7. Accuracy 97%

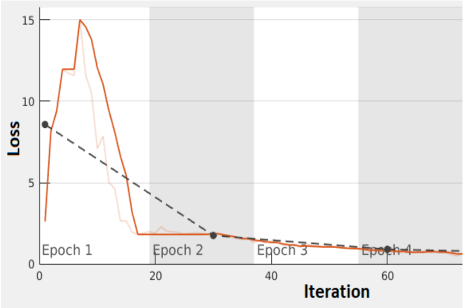

Fig. 8. Loss 80%

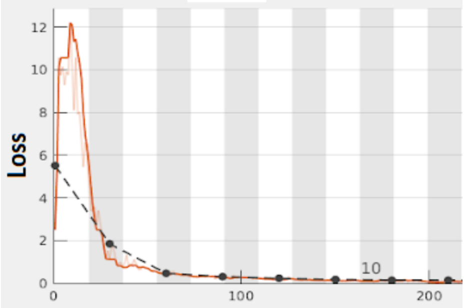

Fig. 9. Loss 97%

Figures 4 and 5 show the performance of the CNN model on original and augmented MSTAR dataset respectively. A clear superiority can be observed in the performance of CNN when it is trained on the augmented dataset. Figures 6 and 7 show the accuracy results for proposed CNN models using different learning rates and epochs. Learning rates are 0.01 and 0.005;

epochs are 10 and 40. Figures 8 and 9 show the cross entropy loss for proposed CNN models using different learning rates and epochs. Learing rates are 0.01 and 0.005; epochs are 10 and 40.

C. Performance metrics and results

The various performance metrics have been used for the performance evaluation such as: precision, recall, F1, specificity, ROC and Geometric mean. Tables I and III details the comparative analysis of the performance results for the different classes with CNN models along with the state-of-art. In [19], Principle Components Analysis (PCA), Independent Components Analysis (ICA), Hu Moments are used as feature extractors for Linear (LDC), Quadratic (QDC), K-nearest Neighbour (K-NN), and Support Vector Machine (SVM) classifiers. The top performance is observed in the case of 3 Nearest Neighbour + PCA Feature extractor. In [8], SVM classifier is combined with a hybrid range and azimuth profiles feature extractor to obtain an accuracy of 90.7%. The performance of our CNN image classifier is better than the traditional machine learning models explored in these papers.

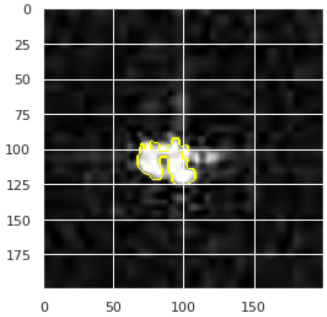

Fig. 10. LIME XAI results for fewer parameters for coarse explainability analysis

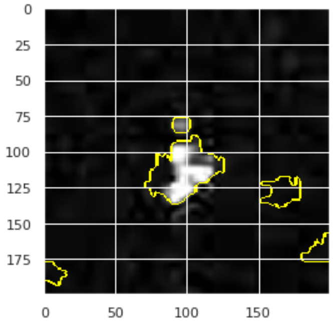

Fig. 11. LIME XAI with more parameters for detailed granular analysis

The table IV shows the performance metrics for all 8 classes for testing.

There were total of 4939 images in MSTAR dataset before augmentation and 6150 images after augmentation as shown in table II with number of classes being eight. The training, validation and test image datset ratio is 64: 16: 20 in case of augmented and original dataset.

D. XAI model - LIME

The predictions made by an ML model can be accepted or rejected depending on the reasoning behind them. A model and the decisions it makes can be trusted when the prior human knowledge about the application domain coincides with the reasoning behind the model's decision. This comparison can only be made if we understand this reasoning. We use LIME as a method to explain models by presenting the representative individual predictions and their explanations in a non-redundant way. This is achieved by displaying visual descriptions that provide a qualitative understanding of the relationship between the instance's elements and the model's prediction. Interpretability is one of the essential criteria for explaining the model's reasoning. This requirement further implies that explanations should be easy to understand and should take the limitations of the user under consideration. In the case of image classification, hundreds or thousands of features significantly contribute to a prediction. It is not reasonable to expect any user to understand the reasoning behind predictions, even if they can inspect individual weights. Interpretable explanations require the use of a representation that is understandable to humans, regardless of the actual features used by the model. For image classification, we use binary vector representation that indicates the "presence" or "absence" of a bordering patch of similar pixels (a super-pixel), while the classifier interprets the image as a tensor with three color channels per pixel. The second fundamental criteria, for the task of explanation, is local fidelity. Often an explanation can't be completely trusted unless it is the complete explanation of the model itself. For an analysis to be significant it must at least be locally faithful, i.e. it must correspond to how the model behaves in the neighborhood of the instance being predicted. The overall goal of LIME is to identify an interpretable model over the interpretable representation that is locally faithful to the classifier. For the image classification task in our paper, we use sparse linear explanations for image classifiers [18]. It provides explanations for targets in the images by highlight the super-pixels with the positive weight towards a specific class as they give intuition as to why the model would think that class may be present. The explanations provided by LIME are depicted in Fig. 10 and 11 in the form of highlighted boundaries around the important features of the images which contributed in making the decisions by black box model. Fig. 10 provides a course analysis by taking into account the main features used for classification of that image whereas Figure Fig. 11 is a granular analysis of the features used for the decision making process.

V. Conclusion

CNN is capable of extracting complex features from images automatically that are intuitively incomprehensible to the human subjective vision. The performance accuracy proposed

TABLE III
COMPARATIVE ANALYSIS OF TRADITIONAL AND THE STATE-OF-ART ON MOVING AND STATIONARY TARGET ACQUISITION AND RECOGNITION (MSTAR) DATASET

Sr.	Reference	Technique	Accuracy
1	Y. Yang [19]	Arch. 1: Linear (LDC)/ Quadratic (QDC) + PCA Feature extractor Arch. 2: 3 Nearest Neighbour + PCA Feature extractor Arch. 3: SVM + PCA Feature extractor Arch. 4: Linear (LDC) + ICA Feature extractor Arch. 5: Quadratic (QDC) + ICA Feature extractor Arch. 6: 3 Nearest Neighbour + ICA Feature extractor Arch. 7: SVM + ICA Feature extractor Arch. 8: Linear (LDC)/Quadratic (QDC) + Hu Feature extractor Arch. 9: 3 Nearest Neighbour + Hu Feature extractor Arch. 10: SVM + Hu Feature extractor	Arch. 1: <50% Arch 2: 98.67% Arch 3: 96.61% Arch 4: 83.06% Arch 5: <50% Arch 6: 95.14% Arch 7: 55.82% Arch 8: <50% Arch 9: 76.85% Arch 10: 73.69%
2	I. M. Gorovyi [8]	SVM + fusion of range and azimuth profiles feature extractor	90.07%
3	Z. Lin [7]	Ensemble; Models: 2 CHU	99.09%
4	H. Furukawa [4]	CNN; Layers: 17 Conv, 1 FC, Based on ResNet-18 [5]	99.56%
5	**Proposed**	**CNN: Layers: 3 Conv + ReLu + Batch Norm, 2MaxPool, 1 flattened, 1 Dense + Softmax**	**98.78** %

TABLE IV
PERFORMANCE RESULTS FOR EACH OF 8 CLASSES

Classes	Prec	Recall	F1	Spec	ROC	GM
2S1	0.94	0.97	0.96	0.99	0.98	0.98
BRDM-2	1.00	0.99	1.00	1.00	1.00	1.00
BTR-60	1.00	1.00	1.00	1.00	1.00	1.00
D7	0.99	1.00	1.00	1.00	1.00	1.00
SLICY	1.00	1.00	1.00	1.00	1.00	1.00
T62	0.98	0.98	0.98	1.00	0.99	0.99
ZIL131	0.99	0.98	0.99	1.00	0.99	0.99
ZSU-23/4	0.98	0.95	0.96	1.00	0.97	0.97

CNN model on the MSTAR dataset is 98.78%. Often the users doubts the model's decisions, especially if the underlying judgement is non-obvious and critical. Therefore XAI is a ideal choice for explaining the test decisions of the model using visual representation. Further improvement in the performance of the model can be achieved by using prominent and diverse datasets which are not for public use. Future plan is to incorporate *Convolutional Auto-encoders (CAE)*, which is an unsupervised method for hierarchical feature extraction. Instead of random initialization of the CNN features, CAE can be used to obtain good initializations for the CNN model. CAE has been used in [6] and are capable of improving the classification accuracy.

REFERENCES

[1] S. Bengio, L. Deng, H. Larochelle, H. Lee, and R. Salakhutdinov, “Guest editors' introduction: Special section on learning deep architectures,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1795–1797, 2013.

[2] “SAR image remote sensing.” https://raweb.inria.fr/rapportsactivite/RA2011/ariana/uid42.html, 2011. [Online; accessed 26 Jan.2020].

[3] C. Coman *et al.*, “A deep learning sar target classification experiment on mstar dataset,” in *2018 19th International Radar Symposium (IRS)*, pp. 1–6, IEEE, 2018.

[4] H. Furukawa, “Deep learning for target classification from sar imagery: Data augmentation and translation invariance,” *arXiv preprint arXiv:1708.07920*, 2017.

[5] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[6] S. Zaied, A. Toumi, and A. Khenchaf, “Target classification using convolutional deep learning and auto-encoder models,” in *2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 1–6, IEEE, 2018.

[7] Z. Lin, K. Ji, M. Kang, X. Leng, and H. Zou, “Deep convolutional highway unit network for sar target classification with limited labeled training data,” *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 1091–1095, 2017.

[8] I. M. Gorovyi and D. S. Sharapov, “Efficient object classification and recognition in sar imagery,” in *2017 18th International Radar Symposium (IRS)*, pp. 1–7, IEEE, 2017.

[9] R. Min, H. Lan, Z. Cao, and Z. Cui, “A gradually distilled cnn for sar target recognition,” *IEEE Access*, vol. 7, pp. 42190–42200, 2019.

[10] R. Chakraborty, Y. Xing, and S. Yu, “Surreal: Complex-valued deep learning as principled transformations on a rotational lie group,” *arXiv preprint arXiv:1910.11334*, 2019.

[11] Q. Liu, S. Li, S. Mei, R. Jiang, and J. Li, “Feature learning for sar images using convolutional neural network,” in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 7003–7006, IEEE, 2018.

[12] M. Heiligers and A. Huizing, “On the importance of visual explanation and segmentation for sar atr using deep learning,” in *2018 IEEE Radar Conference (RadarConf18)*, pp. 0394–0399, IEEE, 2018.

[13] T. D. Ross, S. W. Worrell, V. J. Velten, J. C. Mossing, and M. L. Bryant, “Standard sar atr evaluation experiments using the mstar public release data set,” in *Algorithms for Synthetic Aperture Radar Imagery V*, vol. 3370, pp. 566–573, International Society for Optics and Photonics, 1998.

[14] A. Malhi, T. Kampik, H. Pannu, M. Madhikermi, and K. Främling, “Explaining machine learning-based classifications of in-vivo gastral images,” in *2019 Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1–7, IEEE, 2019.

[15] M. Madhikermi, A. K. Malhi, and K. Främling, “Explainable artificial intelligence based heat recycler fault detection in air handling unit,” in *International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems*, pp. 110–125, Springer, 2019.

[16] H. Kusetogullari, A. Yavariabdi, A. Cheddad, H. Grahn, and J. Hall, “Ardis: a swedish historical handwritten digit dataset,” *Neural Computing and Applications*, pp. 1–14, 2019.

[17] K. Du, Y. Deng, R. Wang, T. Zhao, and N. Li, “Sar atr based on displacement-and rotation-insensitive cnn,” *Remote Sensing Letters*, vol. 7, no. 9, pp. 895–904, 2016.

[18] M. T. Ribeiro, S. Singh, and C. Guestrin, “" why should i trust you?" explaining the predictions of any classifier,” in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144, 2016.

[19] Y. Yang, Y. Qiu, and C. Lu, “Automatic target classification-experiments on the mstar sar images,” in *Sixth International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing and First ACIS International Workshop on Self-Assembling Wireless Network*, pp. 2–7, IEEE, 2005.