International Conference on Advances in Materials, Computing and Communication Technologies
In Association with International Journal of Scientific Research in Science and Technology
Volume 9 | Issue 1 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)

Efficient Explainable Deep Learning Technique for COVID-19 Diagnosis Based On Computed Tomography Scan Images of Lungs

M. Madhavi, Dr. P. Supraja
SRM Institute of Science and Technology, Kattankulathur, Chengalpet, Tamil Nadu, India

ABSTRACT

The entire human race is currently facing a huge disruption of everyday life due to the rapid spread of the novel Corona Virus disease 2019 (COVID-19). It is essential to develop a tool or model for fast diagnosis of the disease which is pandemic and also the model should be able to justify the result for trustworthy in the field of medicine. Machine learning (ML) and Deep Learning (DL) models play a vital role in identifying COVID-19 patients by visually analyzing their Computed Tomography (CT) scan images. In this paper, few publicly available convolutional neural network models (CNN) were analyzed to classify the CT scan images of lungs into two classes, COVID-19 positive and negative cases. In addition to that, Local Interpretable Model-agnostic Explanation (LIME) framework is used as an explanation technique for interpretability. The pixel of relevancy responsible for the outcome of classification is visually explained through LIME technique which gives trustworthiness in the field of healthcare.

Key words: Deep Learning, COVID-19, Local Interpretable Model-agnostic Explanation, Convolutional Neural Network.

I. INTRODUCTION

The covid disease (COVID-19) pestilence [1] was started in Wuhan, a city in China, has promptly spread around various countries all over the world with huge number of cases. The clinical indications of the illness typically start after not exactly seven days, comprising of fever, hack, nasal blockage, weariness and different indications of upper respiratory lot contaminations. The contamination can advance to extreme sickness with dyspnoea and serious chest indications relating to pneumonia in around 75% of patients, as observed by CT scan and X-ray images. Various procedures would be exceptionally essential to manage the current scenario and these integrate computational demonstrating, authentic gadgets, and quantitative examinations to control the spread similarly as the quick advancement of different treatment. Deep learning is quickly turning into the best in class, prompting upgraded execution in different clinical applications. These techniques are effective in image recognition, identification of anatomical and cell structures, segmentation of tissues, computer aided diagnosis, forecasting and so on [2]. Deep learning models, specifically convolutional networks, have quickly become a system of decision

for investigating clinical pictures [3]. Customized CNN framework can consequently and productively take in the intrinsic region of the images highlighted from clinical pictures that are generally appropriate for the classification purpose [4]. Deep learning models are the bleeding edge of the advancement of Machine Learning calculations, because of their nested non-linear structure, these incredible models have been for the most part considered as "black boxes", not giving any data about what precisely causes them to show up at their expectations. Since in numerous applications, e.g., in the clinical area, such absence of straightforwardness might be not worthy, the improvement of strategies for imagining, clarifying and deciphering deep learning models has as of late pulled in expanding consideration. At the point when the AI suggests an output, the decision makers would have to comprehend the fundamental explanation. The exploration progress in the area of Explainable Artificial Intelligence (XAI) has been quickly increasing. Local Interpretable Model-agnostic Explanations (LIME) [5] is an intermediary strategy that uses the learning of local substitute models to clarify deep neural organizations. LIME gives a patient-explicit explanation to a given characterization, subsequently upgrading the opportunities for any perplexing classifier to fill in as a wellbeing help inside a clinical setting.

In this paper, our objective is to explore the different publicly available CNN models which automatically diagnose the existence of COVID-19 in Computed Tomography (CT) scan images. Accuracy for each model is compared for COVID-19 positive and negative cases. Specifically, we used LIME explanation model for prediction explanations.

II. RELATED WORKS

Y. Xu et al., [6] utilized Multiple Instance Learning (MIL) structure in order to train the features for deep learning classification. Ahmad, M. A., et al., [7] investigates the scene of late advances to address the difficulties in model interpretability in the field of medical services and furthermore depict how one would approach in choosing the privilege interpretable AI algorithm for a given issue in medical care. Amine Amyar et al., [8] proposed to perform various tasks learning way to deal with the classification of COVID-19 positive and negative cases from CT scan images and sectioning the areas of attention. A 22-layered CNN architecture was proposed by Emtiaz Hussain et al., [9] for COVID-19 classification in deep learning based identification by utilizing X-ray images of chest. Adel Oulefki et al., [10] introduced an automated tool for division, segmentation and estimation for COVID-19 lung infection utilizing CT scan images of chest. This technique can distinguish abnormal areas with low-intensity contrast among sores and solid tissues. An exhaustive overview to give direction in the area of XAI and the appraisal of quantitative assessment measurements for reasonableness is proposed by Aniek F et al., [11]. Samek, W [12] proposed two ways to deal with clarify forecasts of deep learning models, one technique which processes the sensitivity of the interpretability regarding changes in the input and the other methodology is Layerwise Relevance Propagation (LRP) in which genuinely deteriorates the choice as far as the information factors. Jean-Baptiste Lamy proposed a visual and logical Case Based Reasoning framework joining quantitative and subjective methodologies [13]. Garreau, D et al., [14] inferred closed form structure for the interpretable model coefficients when the capacity to clarify is linear. Munkhdalai, L et al., [15] explore the high level neural organization methodology and its justification for credit scoring. The method utilized is LIME to decipher the black box and confirm its dependability by looking at a more interpretable calculated model. Hung, S.C et al., [16] developed another convolutional model called RSSCNet (remote sensing scene classification network), with high speculation ability was planned and LIME (local interpretable model agnostic explanation) calculation

was applied to improve the outcomes in situations where the model made wrong forecasts. Cian, D et al., [17] run two strategies for explanation, to be specific LIME and Gradient based Class Activation Mapping (Grad-CAM), on a convolutional neural organization prepared to mark images with the LEGO blocks that are noticeable in them. Pavan Rajkumar Magesh et al., [18] proposed an AI model that precisely orders any given DaTSCAN as having Parkinson's infection or not, as well as giving a conceivable motivation to the expectation using LIME technique.

III. METHODS AND MATERIALS

A. Convolutional Neural Network

A ConvNet/CNN is Convolutional Neural Network is a Deep Learning technique that takes information from input image, allot significance to loads and inclinations which is nothing but weights and bias to different features in the images and have the option to segregate one from the other. When comparing with the other classification algorithms, the pre-processing steps that to be followed in a ConvNet is much lower. ConvNets is capable of learning filters or characteristics automatically. It has the ability to capture spatial and temporal dependencies in an image with the help of significant channels. Some commonly used architectures for convolutional networks are LeNet, AlexNet, VGG-16, Inception (GoogleNet), ResNet, ResNeXt, DenseNet etc., In this paper, we have used some of the publicly available CNN models for classification of COVID-19 using Computed Tomography scan images.

AlexNet :

AlexNet was the first convolutional network which used GPU to boost performance. The architecture comprises of 5 convolutional layers, 3 max-pooling layers, 2 normalization layers, 1 softmax layer and 2 fully connected layers. All convolutional layer is composed of convolutional channels and a nonlinear activation function ReLU. Max pooling is done through the pooling layers. Due to the existence of fully connected layers the input size is fixed and in general, this type of CNN model has the parameters of around 60 million.

VGG –16 :

VGG-16 is a less complex engineering model, since it is not utilizing a large number of hyper parameters. It generally utilizes 3 x 3 channels with the stride of one in convolution layer and uses similar padding in pooling layers 2 x 2 with the stride of two.

GoogleNet :

GoogLeNet model is otherwise called as Inception Module. It accomplished an error rate of 6.67% which is considered as a top 5 among all architectures and goes deeper in parallel paths with various open field sizes. A 22 layer deep architecture is used in this model whereas the hyper parameters in AlexNet were 60 million whereas in GoogleNet it is reduced to 4 million.

ResNet :

Residual Neural Network architecture presented an idea called "skip connections" to take care of the issue of the vanishing or detonating slope. The skip connection omits training from a few layers and interfaces straightforwardly to the output. Ordinarily, the information grid computes in two linear transformations with the activation function ReLU. In Residual organization, it straightforwardly duplicates the information lattice to the subsequent change yield and entirety the yield in the final ReLU function.

InceptionV3 :

Inception-V3 is a convolutional neural architecture design from the family of Inception where few upgrades are done including the usage of smoothing label, factorized 7 x 7 convolutions, and the utilization of a helper classifier to proliferate name data lower down the organization (alongside the utilization of bunch standardization for layers in the

sidehead). InceptionV3 is an extensively used image recognition model that has been seemed to achieve more important than 78.1% precision on the ImageNet dataset. The genuine model is involved symmetric and disproportionate design blocks, including convolutions, ordinary pooling, max pooling, concats, dropouts, and fully connected (FC) layers. Batchnorm is utilized broadly all through the model and applied to enactment inputs. Misfortune is figured through Softmax.

These different Convolutional network architectures were used to train the COVID-19 image dataset for binary classification between the infected and non-infected classes. Training performance and classification accuracy between these models were compared for our classification study.

B. Local Interpretable Model-agnostic Explanation

Local Interpretable Model-agnostic Explanation (LIME) gives a quick and generally straightforward technique for locally clarifying discovery models. The LIME algorithm functions as follows: For a given data point, self-assertively perturb its highlights more than once. In plain information, this includes adding an unobtrusive amount of commotion to each element. Get expectations for each perturbed data. This helps us with building up a nearby picture of the decision surface by at that point. Use conjectures to handle an expected direct "clarification model" using assumptions. Coefficients of the immediate model are used as explanations. The LIME python library gives interfaces to clarifying models based on plain (TabularExplainer), picture (LimeImageExplainer), and text information (LimeTextExplainer). In this methodology, a proxy model is utilized to estimate the basic social of the deep neural organization. The deep neural organization is examined and the proxy model is prepared dependent on the forecast yields of the deep learning model. Diverse stage of samples is created and another dataset is built dependent on the produced tests and the relating expectations of the deep neural network. At that point, an interpretable model (i.e., substitute model) is utilized and is prepared by the produced dataset. Accordingly, the preparation cycle can be formed as follows:

$$I(x) = \underset{g \in G}{\operatorname{argmin}}\ L(f, g, \theta) + \Gamma(g) \quad (1)$$

where I(x) encodes the clarification for example x by means of the ideal capacity g in the conceivable arrangement of capacity G. f addresses the deep neural organization, Γ( ) distinguishes the intricacy of the capacity g to be utilized in the logic interaction, and L is the misfortune work which estimates the similitude of substitute model and the first model (i.e., here the deep neural organization). In this investigation, we influence edge relapse models as the proxy models. The work stream of LIME engineering is given in Figure1.

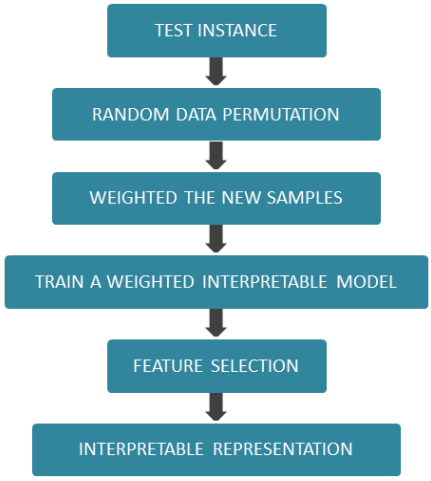

Fig. 1. Flow diagram of LIME working progress

IV. EXPERIMENTAL SETUP

This section comprises of experimental setup for identification of COVID-19 with explainable deep learning technique. In this work, we have utilized pre-prepared information to improve the analytic exhibition utilizing move learning procedures and thought about the presentation of various CNN structures. Here we have given detailed description about the dataset used and the LIME explainability model.

A. Dataset:

In this paper, the CT scan images of lungs by Xingyi Yang et al., [19]are considered for the COVID-19 positive cases classification.COVID-19 data are sensitive, people are finding very difficult in data collection. Almost all data are available as open, getting the entire data set for scientific research is very difficult. Xingyi Yang et. al., 2020 collected the highly reliable CT scan images of 349 COVID-19 positive cases and 463 Non-COVID cases (classes) for their research purpose. In a human body, the major target region of COVID-19 virus is Lung. The description of Lung can be visualized using the chest CT scans. CT scan images gathered for the computational research by Xingyi Yang et. al., 2020, are considered for the classification process. The authors collected the images in two categories, COVID and Non-COVID images. The sample set of images of both positive and negative cases images are shown in Fig 2(a) and 2(b) respectively.

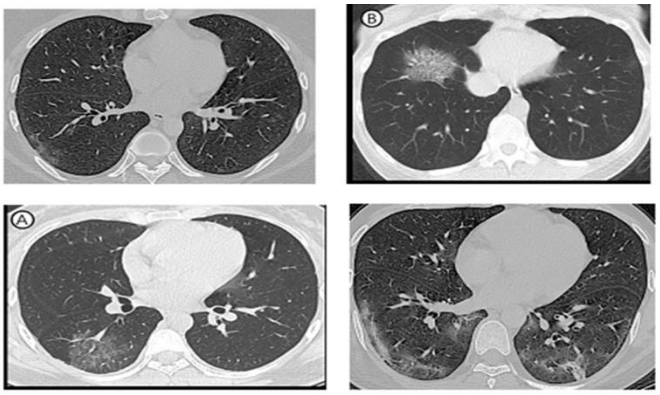

Fig 2(a) Sample CT scan images of COVID-19

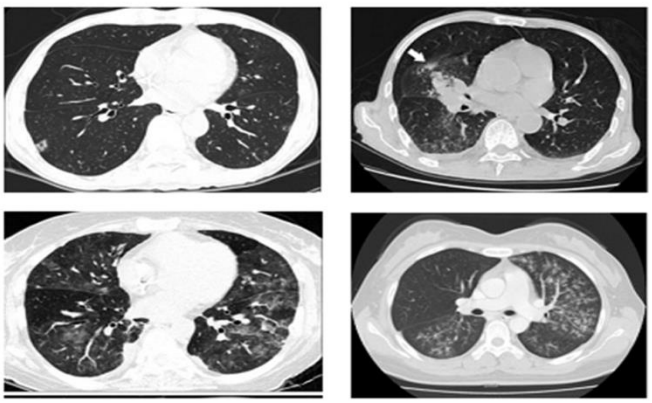

Fig 2(b) Sample CT scan images of Non COVID-19

In this process, the basic transformation operations such as rotation, scaling and shearing are applied on the images. The augmentation process is applied for both training and testing data set.

B. Explainable CNN model:

The proposed network framework is shown in Figure 3. As shown in the figure, the CT scan images of the subjects were collected and stored in the database. These images were categorized as COVID-19 positive and negative cases. Both the positive and negative cases were undergone a deep neural network for automatic prediction. Here we have used five different pre-trained models (AlexNet, VGG-16, GoogleNet, ResNet and InceptionV3) for training and testing process. These models were used for binary classification of positive and negative cases.

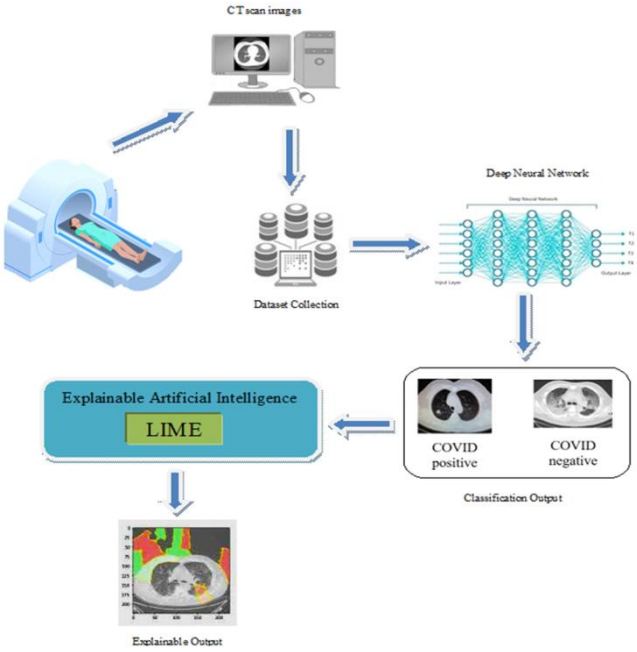

Fig. 3.Schematic Representation of the explainable CNN with LIME

We utilize local interpretable model-agnostic explanations for the justification of the expected output obtained from our model. The processed model weights were the contribution to LIME explainability algorithm which performs annoyances on the highlights over a bunch of anticipated

examples on the datasets, which decipher the weights came about in the wake of preparing these models.

V. RESULTS AND DISCUSSIONS

In this section, the results obtained from five different pre-trained CNN models were compared. Here we discuss about the training performance of each model in which the training loss, validation loss and validation accuracy is compared in Table I. Similarly the classification performances of all five models were discussed in Table II where the precision, sensitivity, specificity, F1-score, Accuracy and AUC.

TABLE I: TRAINING PERFORMANCE

	Training Loss	Validation Loss	Validation Accuracy
AlexNet	0.24	0.004	99.04
VGG-16	0.1727	0.187	96.12
GoogleNet	0.0026	0.0473	98.58
ResNet	0.0017	0.0268	98.11
InceptionV3	0.3458	0.1402	95.17

When the accuracy looks high, the loss would be less in the specific level. Validation accuracy is high in Ale Net followed by GoogleNet and ResNetcompared with the remaining pre-trained models.

TABLE II: CLASSIFICATION RESULT COMPARISON

	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score	Accuracy (%)	AUC
AlexNet	96.73	98.34	96.63	0.9748	97.50	0.9638
VGG-16	95.06	96.63	95.02	0.9582	95.81	0.9479
GoogleNet	96.63	96.63	96.63	0.9663	96.63	0.9698
ResNet	96.79	99.94	97.81	0.9831	98.3	0.9839
InceptionV3	96.34	88.45	96.63	0.9214	92.48	0.9341

ResNet gives highest diagnostic performance among all the five pre-trained models with the sensitivity of 99.94%, specificity 97.81%, and accuracy 98.3% and the performance is excellent in the testing data set when compared with all other four pre-trained models. In the external testing data set, ResNet model shows the detection accuracy as highest (98.3%), followed by AlexNet, GoogleNet, VGG16 and InceptionV3 (97.5%, 96.63%, 95.8% and 92.48%, respectively).

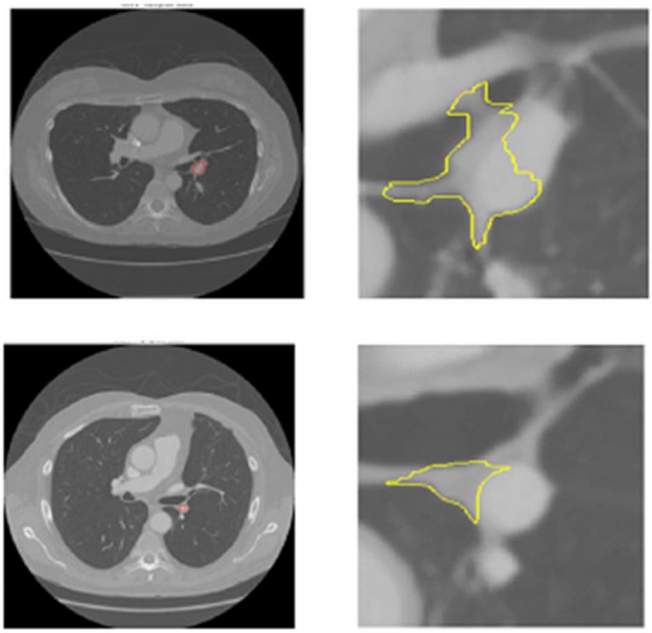

Fig. 4. LIME pixel of relevancy with classification outcome

The decision for the diagnosis can be explained with the reason by our proposed method with the indication of highlighting the important regions on the image. The important regions are highlighted in the image by using LIME technique. The pixel of relevancy responsible for the outcome of classification is represented in Fig. 4. The results shows that the proposed method with the LIME explainer can explain the diagnostic decision more effectively in the deep neural network.

VI. CONCLUSION

In this work, a deep neural network based classifier is analyzed and a quantitative assessment showed high

precision for programmed conclusion of COVID-19. Pre-trained deep learning models used in this investigation could be utilized for early screening of covid to analyze COVID-19 and non COVID-19 positive and negative cases from computed tomography scan images dataset. We have utilized five diverse pre-trained models made out of neural organization based structures AlexNet, VGG-16, GoogleNet, ResNet and InceptionV3 which can be embraced to recognize or characterize COVID-19 cases. These models were prepared for binary classification and the preparation execution and accuracy comparison of every one of these models were considered. We further show that our expectations are interpretable utilizing LIME structure. In future, we plan to experiment on a large dataset and more reasonable procedures for multimodal support of the translation.

VII. REFERENCES

[1]. Velavan, T. P., & Meyer, C. G. (2020). The COVID-19 epidemic. Tropical medicine & international health, 25(3), 278.

[2]. Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. Annual review of biomedical engineering, 19, 221-248.

[3]. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88.

[4]. Li, Qing, et al. "Medical image classification with convolutional neural network." 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV). IEEE, 2014.

[5]. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Model-agnostic interpretability of machine learning. arXiv preprint arXiv:1606.05386.

[6]. Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai and E. I. Chang, "Deep learning of feature representation with multiple instance learning for medical image analysis," 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 2014, pp. 1626-1630, doi: 10.1109/ICASSP.2014.6853873.

[7]. Ahmad, M. A., Eckert, C., & Teredesai, A. (2018, August). Interpretable machine learning in healthcare. In Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics (pp. 559-560).

[8]. Amine Amyar, Romain Modzelewski, Hua Li, Su Ruan, Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation, Computers in Biology and Medicine, Volume 126, 2020, 104037, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2020.104037.

[9]. Emtiaz Hussain, Mahmudul Hasan, MdAnisur Rahman, Ickjai Lee, TasmiTamanna, Mohammad ZavidParvez, CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images, Chaos, Solitons & Fractals, 2020, 110495, ISSN 0960-0779, https://doi.org/10.1016/j.chaos.2020.110495.

[10]. Adel Oulefki, SosAgaian, ThaweesakTrongtirakul, AzzeddineKassahLaouar, Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images, Pattern Recognition, 2020, 107747, ISSN 0031-3203, https://doi.org/10.1016/j.patcog.2020.107747.

[11]. Aniek F. Markus, Jan A. Kors, Peter R. Rijnbeek, The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies, Journal of Biomedical Informatics, 2020, 103655, ISSN 1532-0464, https://doi.org/10.1016/j.jbi.2020.103655.

(http://www.sciencedirect.com/science/article/pii/S1532046420302835)

[12]. Samek, W., Wiegand, T., & Müller, K. R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296.

[13]. Jean-Baptiste Lamy, BoomadeviSekar, Gilles Guezennec, Jacques Bouaud, Brigitte Séroussi, Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach, Artificial Intelligence in Medicine, Volume 94, 2019, Pages 42-53, ISSN 0933-3657, https://doi.org/10.1016/j.artmed.2019.01.001. (http://www.sciencedirect.com/science/article/pii/S0933365718304846)

[14]. Garreau, D., &Luxburg, U. (2020, June). Explaining the explainer: A first theoretical analysis of LIME. In International Conference on Artificial Intelligence and Statistics (pp. 1287-1296). PMLR.

[15]. Munkhdalai, L., Wang, L., Park, H. W., &Ryu, K. H. (2019, April). Advanced neural network approach, its explanation with lime for credit scoring application. In Asian Conference on Intelligent Information and Database Systems (pp. 407-419). Springer, Cham.

[16]. Hung, S. C., Wu, H. C., & Tseng, M. H. (2020). Remote sensing scene classification and explanation using RSSCNet and LIME. Applied Sciences, 10(18), 6151.

[17]. Cian, D., van Gemert, J., &Lengyel, A. (2020). Evaluating the performance of the LIME and Grad-CAM explanation methods on a LEGO multi-label image classification task. arXiv preprint arXiv:2008.01584.

[18]. Pavan Rajkumar Magesh, Richard Delwin Myloth, Rijo Jackson Tom, An Explainable Machine Learning Model for Early Detection of Parkinson's Disease using LIME on DaTSCAN Imagery, Computers in Biology and Medicine, Volume 126, 2020, 104041, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2020.104041. (https://www.sciencedirect.com/science/article/pii/S0010482520303723).

[19]. Xingyi Yang, Xuehai He, Jinyu Zhao, Yichen Zhang, Shanghang Zhang, and PengtaoXie. "COVID-CT-dataset: a CT scan dataset about COVID-19." ArXiv e-prints (2020): arXiv-2003.