Analysis of Radiomics Features of Rib Fracture X-ray extracted from Pyradiomics and selected by Machine Learning Algorithms

Classification of Rib Fracture on Chest X-ray with Radiomic feature extraction and Machine Learning Algorithms

Radiomic Features Analysis in Rib Fractures on Chest X-ray with Machine Learning Classification methods

Classification of Rib Fracture on Chest X-ray: using Radiomic features and Machine Learning Algorithms

Abstract

Rib fractures are one of the most commonly occurring fractures and indeed one of the most dangerous fractures. A chest X-ray is the most accessible method to identify rib fracture visually, but there is a limitation that some types of fractures are not easily detectible with human visons. Advancing and developing a computer-aided diagnosis system in the classification of rib fracture on X-ray images can help emergent patients with chest traumas. We extracted Radiomics features of rib fracture in chest X-ray using Pyradiomics and used feature selection method through four machine learning algorithms: logistic regression, random forest classifier, extreme gradient boosting, and support vector machine. We used 253 patient's chest X-rays and 360 ROIs as data samples to extract image features that were used to train and test machine learnings. 8 out of 115 features were selected during feature selection, and they measure the texture and asymmetry characteristic of an image in different ways. The highest area under the receiver operating characteristic curve was 0.957 (95% CI from 0.900 to 0.987) from a support vector machine classifier. Our analysis of Radiomics features can contribute to the future development of computer diagnosis system in rib fracture to help doctors and patients to identify the broken ribs in early pace.

Introduction

Rib fractures are very common thorax injury in daily life (25%) [1] that is mostly caused by chest trauma such as from a fall, motor vehicle accident, or impact during contact sports [2]. A broken rib may cause serious associated injuries even to death. It is already clearly reported that among 711 people who had rib fractures, 84 (12%) died, 670 (94%) had associated injuries, 272 (34%) had a hemothorax or pneumothorax, and 187 (26%) had a lung contusion. [3] Considering their mortality, it is seriously critical to identify a patient's rib fracture in early pace.

There are various ways to identify rib fractures. They can be easily suspected by the patient's history and pain clinically, but it is also important to identify the exact location of the broken part of the rib for accurate diagnosis and rapid recovery [4]. The precise location can be recognized by radiographic technologies such as chest X-ray or CT scan. CT scan is much accurate and easy to see fractured bones as it is sometimes called an upgrade version of an x-ray [5]. However, due to the high expense of CT, the machines are often not supplied in rural and small hospitals. Moreover, the average effective radiation dose, which can increase the risk of getting cancer in a lifetime, from chest CT (7 mSv) is 70 times higher than from chest X-ray (0.1 mSv). For these reasons, identification of rib fracture in X-ray remains a significant issue we should consider.

Despite their vital significance, depicting rib fracture in an X-ray is not accurate that 25% of them do not show on it with human eyes [7]. A recent study by J. F. Griffith has examined that only 8 out of 83 (10%) rib fractures are detected in radiographs [8]. Relatively minor fractures such as undisplaced fractures, impacted fractures, or stress fractures are not easily visible and can yield a false negative result [9]. To overcome these limits on recognizing rib fracture in X-ray, a computer-aided method with image features can be useful to assist clinicians [10]. Especially, machine learning applying Radiomics features can be used to reduce errors and increase accuracy. The term "Radiomics" refers to the extraction of magnificent image features from radiographic data using data-characterization algorithms [11]. Radiomics are widely used in medical imaging and high-potential as a study of lung nodule classification using Radiomic features suggests that accuracy was 84% with the sensitivity of 92.85% and the specificity of 72.73% [12]. However, the classification of rib fractures using Radiomics is not yet studied and researched widely. Analyzing these quantitative data of features will develop a computer-aided diagnosis to classify rib fractures with much higher accuracy compared to naked eyes.

In this study, we aim to analyze these Radiomics features selected by applying machine learning algorithms to classify rib fractures in X-ray images. Pyradiomics and four machine learning algorithms (logistic regression, random forest classifier, extreme gradient boosting, and support vector machine) are applied to extract features from chest X-ray and to find feature importance. We adopted the receiver operating characteristics curve to compare machine learning methods.

Methods

Data

A total of 253 patients' chest X-rays were provided by Boramae Medical Center (IRB number: 10-2019-50) and used as data samples to train as well as test machine learning models. Among them, 128 patients are diagnosed with a rib fracture, and a region of interest (ROI) is identified as a shape of a square containing a fractured bone. The other 125 patients didn't show any reason for rib fracture, and we randomly sampled ROI to describe a non-fractured normal rib.

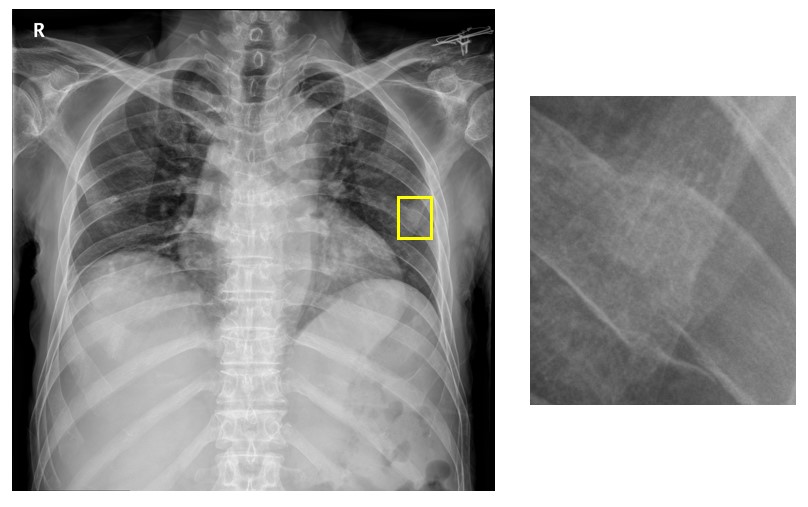


Fig. 1. Original chest X-ray and a cropped ROI image of a broken rib

Image Data Processing

We converted DICOM files of rib fracture X-ray into 8-bit PNG files to work with supporting Python libraries. Then we generated the binary mask files which were segmented by the following ROIs with the same sizes as their counterparts. More than one mask file can be generated from an X-ray since multiple fractures can be found from a patient. Among the total of 360 pairs of X-rays and mask files, 252 samples as training data, and 108 of them as test data are prepared.

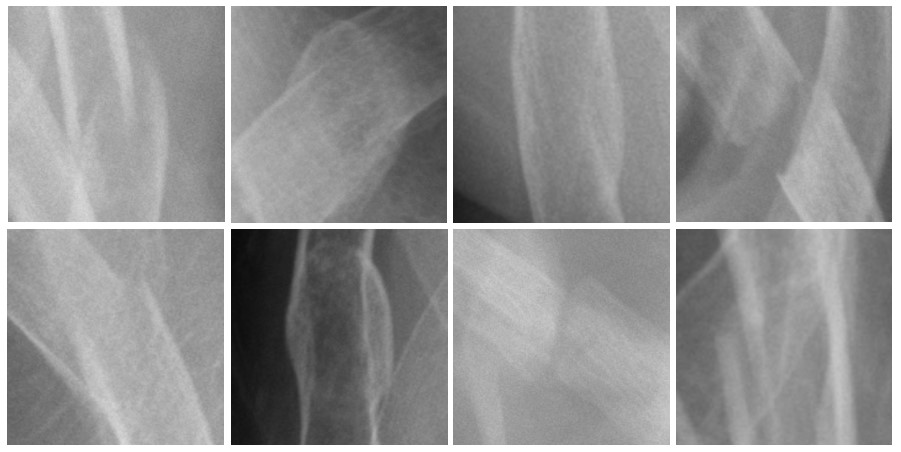


Fig. 4. Various cases of rib fractures on chest X-ray where regions of interest were drawn

Feature Extraction

Pyradiomics, an open-source Python package [ver 3.6.2, OR, Python Software Foundation], was used to extract image features from the processed images. Pyradiomics was developed to extract Radiomics features from medical imaging. There are 7 types of Radomic features used in Pyradiomcis; first-order, gray level co-occurrence matrix (GLCM), gray level size zone matrix (GLSZM), gray level run length matrix (GLRLM), neighboring gray-tone difference matrix (NGTDM), gray level dependence matrix (GLDM). These sets of features without first-order are based on second-order statistics which describes correlations between pairs of pixels in different aspects such as homogeneity or uniformity. To briefly introduce these features, first-order features are in concern of single-pixel rather than any other relationship between pixels of an image. A GLCM represents the number of times the combination of certain two numbers occurring in two neighboring pixels in the image [13]. For a GLSZM, it provides a size of consequence and collinear pixels that share the same gray level intensity, where pixels are considered connected if the distance is 1. Alike to GLSZM, a GLRLM quantifies pixels' length that has the same gray level values, but the pixels only have to be consecutive to be counted [14]. NGTDM compares a gray value of a pixel and an average value of neighboring pixels of it and stores as a form of a matrix. For the last one, GLDM measures the number of connected pixels which are dependent on the center pixel [15].

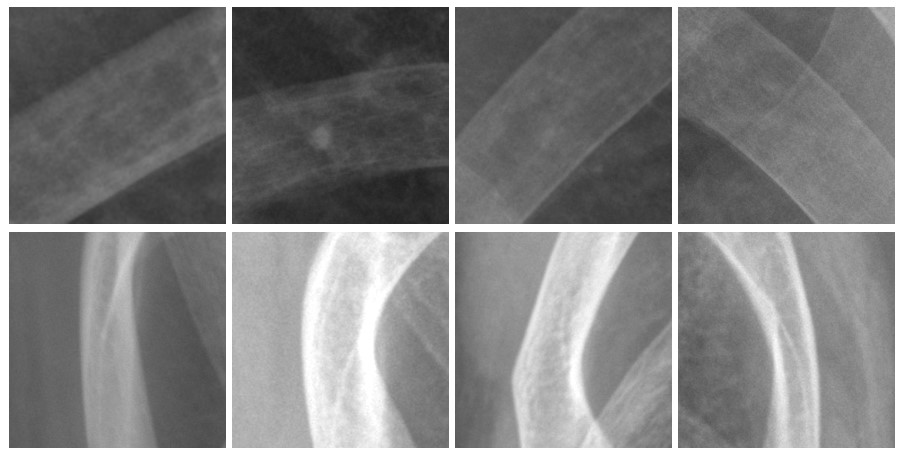


Fig. 5. Cropped chest X-ray images of non-fractured rib where regions of interest were drawn

Feature Selection

Feature selection is an essential step in constructing a machine learning model. By reducing a large number of initial too-many features into a small number to avoid an overfitting problem and to increase accuracy. Overfitting occurs when a part of predictors performs no useful function or when there are too many predictors [16]. It is highly expected that our model will cause an overfitting issue without feature selection because Pyradiomics provides more than a hundred number of features. According to a study done by Jianping Hua and Zixiang Xiong, the ideal feature size for the given 360 sample sizes would be lower than 10 for highly correlated features [17]. Four regression and classification methods are used for feature selection: logistic regression, random forest, extreme gradient boosting, and support vector machine. Sci-kit Learn library was implemented to calculate the feature importance values for four methods. We trimmed features with low importance values to make them lower than 10 and compared models to find which fit the best for this case.

Result

As a result of feature extraction, a total of 115 Radiomics features were extracted from a pair of chest X-ray and mask file. We deleted inappropriate data such as non-integer and redundant values, and 81 features for every 360 samples were left to analyze.

Feature selection is a vital process for a machine learning model to avoid worse classification accuracy. To reduce the number of features lower than 10, we trimmed a feature that had the lowest significance and measured areas under the curves (AUC) for each receiver operating characteristic (ROC) curve repeatedly. After several iterations, the highest mean of AUC was recorded when 16 features are left, but it was still too large to avoid overfitting of features. To solve this problem, we tried the same procedures again by cutting 6 more features and examining which set of features shows the highest AUC.

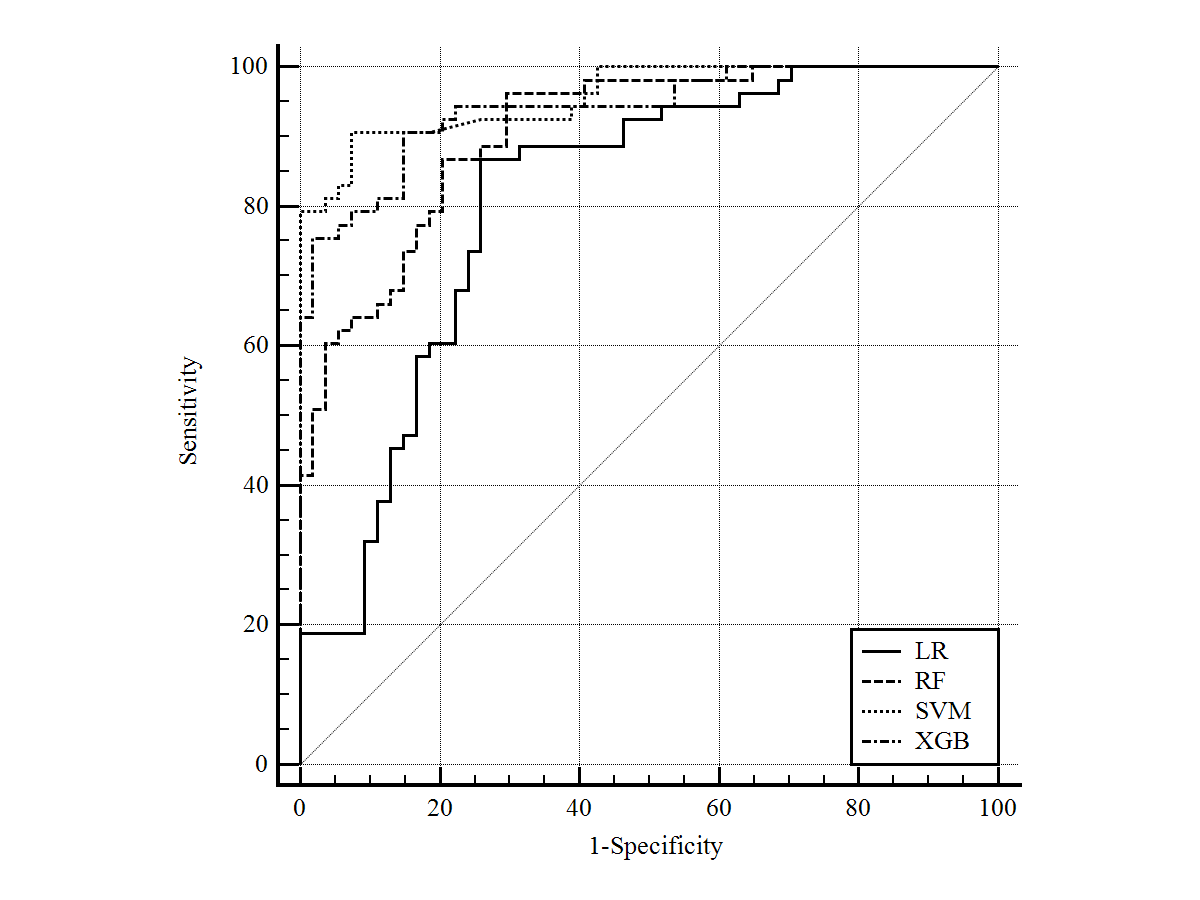


Fig. 2. ROC curves for Logistic Regression, Random Forest, Support Vector Machine, and Extreme Gradient Boost

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Feature | LR | RF | XGB | SVM | Mean |
| GLNU | 0.33 | 0.03 | 0.06 | 0.15 | 0.14 |
| Skewness | 0.25 | 0.04 | 0.08 | 0.19 | 0.14 |
| ZP | 0.27 | 0.02 | 0.05 | 0.19 | 0.13 |
| RMAD | 0.25 | 0.01 | 0.06 | 0.14 | 0.12 |
| CS | 0.03 | 0.10 | 0.11 | 0.18 | 0.11 |
| Complexity | 0.20 | 0.01 | 0.05 | 0.11 | 0.09 |
| Busyness | 0.16 | 0.01 | 0.03 | 0.15 | 0.09 |
| SumSquares | 0.10 | 0.03 | 0.01 | 0.20 | 0.08 |

Fig. 3. Comparison of feature importance of selected features for machine learning algorithms

As a consequence, the highest mean of AUC was earned when the following 8 features are left: GLRLM grey level non-uniformity, first-order skewness, GLSZM zone percentage, first-order robust mean absolute deviation, GLCM cluster shade, NGTDM complexity, NGTDM busyness, and GLCM sum squares. The highest AUC among the four models was 0.957 obtained using a support vector machine (95% CI from 0.900 to 0.987). An extreme gradient boosting (0.94, 95% CI from 0.876 to 0.976), random forest (0.907, 95% CI from 0.835 to 0.955), and logistic regression (0.811, 95% CI from 0.723 to 0.880) were followed next.

To introduce what each feature signifies, GLRLM grey level non-uniformity measures the similarity of the intensity of gray level runs. The skewness of the image means the asymmetry of the distribution of pixel value about the mean. Where higher value means that an image has a more fine texture, the GLSZM zone percentage measures texture of the image by dividing the number of pixels in ROI into the number of zones. Similar to skewness, cluster shade is a measure of homogeneity and uniformity, and the sum of squares is a measure of the distribution of neighboring intensity level pairs about the mean of GLCM. Complexity and busyness literally show how complex and busy the texture of the image is. For the last one, a robust mean absolute deviation is the average distance of all gray run values from the mean value.

Discussion

After operating machine learning algorithms to reduce the number of features to increase accuracy and credibility, 8 features that describe Radiomics of rib fractures in X-rays are left. Based on the result, there is a significant distinction between X-rays of a normal and fractured rib, and it made those 8 features the most important features classifying whether they are fractured or not.

A fractured rib X-ray has a simple and uncomplicated texture compared to a non-fractured. There is a high similarity of gray level's energy in an image of a fractured rib as GLRLM gray level non-uniformity (GLN) means homogeneity in intensity. It is very convincing that an intersection of bones contributed to the difference of gray level uniformity between the groups. As can be seen in figure 5, a non-fractured rib is in the shape of thin, curved, and solid lines. However, ribs in the fracture group are in abnormal shapes due to multiple broken parts and expansion from bone healing. When a bone is fractured, a broken site will be bleeding and blood clotting will happen. Afterward, it will heal itself by replacing from blood clotting to fibrous tissue and cartilage, and from soft tissue into the hard bone until completely remodeled. [18] The expansion of bone results in high uniformity of gray level runs. Besides, a bone is cracked into several parts most of which are figured as overlapped on an X-ray. These overlapped images of bones and expansion of healing bone eventually formulate higher similarity in gray level intensity values. This phenomenon also affects the texture of an image that GLSZM zone percentage, NGTDM complexity, and NGTDM busyness are in concern of. These features evaluate whether an image texture is fine, complex, or busy. On an account of segment and pieces of fractured bones, an X-ray of fractured bone often looks blurry and unclear with image noise, but in figure 5, the non-fractured rib has a very clear and definite outline which means there is a rapid change of intensity between pixels in non-fractured images. The rapid gray level energy change between pixels will make the texture of the image more complex and busier. There is a significant difference in texture between the two groups which makes those four features important in the classification of rib fracture.

It is a widely known fact that a common location of rib fracture is at rib levels 4 to 7 in the lateral and posterolateral segments. A rib fracture does not commonly occur at costal cartilages [19] which is located on the darker background in X-ray because of the lower density of lungs. On the other side, lateral ribs are considered as much brighter in the image with higher density. For this reason, the fractured group has a much higher lightness rating than the normal group with an important relationship to first-order skewness, first-order robust mean absolute deviation, GLCM cluster shade, and GLCM sum squares. A skewness of image tends to be closer to zero when an image appears darker and glossier than when it is lighter and matte, which indicates that the lightness rating of an image has a strong negative dependency on skewness. [20] The rest of the three features also measures the difference of distribution of grey level intensity, which values will be high if an image is significantly skewed. The common occurrence of fracture in a certain location of ribs was a significant factor that influenced skewness and distribution of intensity, and it became one of the features with the strongest importance to classify whether the rib is fractured or not.

A support vector machine classifier showed the best AUC among the four methods we used. While we have an outstanding AUC score of 0.957 of AUC when using the SVM algorithm to classify a rib fracture, more training and testing data samples are needed to improve the credibility of our study. The precision of machine learning algorithms largely depends on the size of data samples, but the SVM tends to show higher accuracy when the sample size is smaller [21]. The AUCs of ROC curves would have been different if we have used a larger quantity of data set. Another weakness of this study can be found that all ROIs are drawn manually instead of automatically. The automated ROI can achieve successful results by improving the objectivity and reproducibility of samples. The manual drawing of ROIs of non-fractured ribs may influence the bias and subjectivity of samples.

As a part of our future work, we would like to develop a detection algorithm for rib fractures to advance this technology useful in real life. We also plan to construct a CAD system not only for a chest X-ray but also for every part of the X-ray to detect bone fractures on the body.

Conclusion

We have extracted Radiomics features from rib fractures in chest X-ray and analyzed those features using machine learning algorithms to find out why some features are especially important and to increase the accuracy and efficiency of machine learning in the classification of rib fracture. A total of eight features are found to be significant because of different gray intensity values and skewness in pixels of the images between the fractured and non-fractured rib radiographs. The performance of the support vector machine was the finest. Our study will be helpful in interpretation as well as identification of rib fracture in chest X-ray by utilizing Radiomics features of it and machine learnings in computer-aided diagnosis systems.

References

[1] Hurley, Maja E., Geoff D. Keye, and Samuel Hamilton. "Is ultrasound really helpful in the detection of rib fractures?." Injury 35, 6, 562-566, 2004.

[2] Broken ribs. (2018, December 01). Retrieved August 03, 2020, from https://www.mayoclinic.org/diseases-conditions/broken-ribs/symptoms-causes/syc-20350763

[3] Ziegler, Daniel W., and Nikhileshwer N. Agarwal. "The morbidity and mortality of rib fractures." The Journal of trauma 37.6 (1994): 975-979.

[4] Kuo, Kevin, and Amie M. Kim. "Rib Fracture." (2019).

[5] C, R. (2019, May 14). Difference Between X-ray and CT scan (with Comparison Chart). Retrieved August 07, 2020, from https://biodifferences.com/difference-between-x-ray-and-ct-scan.html

[6] Publishing, H. (2020, January 29). Radiation risk from medical imaging. Retrieved August 03, 2020, from https://www.health.harvard.edu/cancer/radiation-risk-from-medical-imaging

[7] Abed-Al Nasser Assi, Yasser Nazal. "Rib fracture: Different radiographic projections." Polish journal of radiology 77.4 (2012): 13.

[8] Griffith, J. F., et al. "Sonography compared with radiography in revealing acute rib fracture." AJR. American journal of roentgenology 173.6 (1999): 1603-1609.

[9] Laurin, Sven, Kjell Jonsson, and Ragnar Jonsson. "Low frequency of missed or invisible hip fracture in X-ray examination. The rule-of-2--a simple method for quality assessment." Lakartidningen 101.30-31 (2004): 2423-2425.

[10] Kim, Jaeil, et al. "Quantitative measurement method for possible rib fractures in chest radiographs." Healthcare Informatics Research 19.3 (2013): 196-204.

[11] Lambin, Philippe, et al. "Radiomics: extracting more information from medical images using advanced feature analysis." European journal of cancer 48.4 (2012): 441-446.

[12] Chen, Chia-Hung, et al. "Radiomic features analysis in computed tomography images of lung nodule classification." PloS one 13.2 (2018): e0192002.

[13] Gadelmawla, E. S. "A vision system for surface roughness characterization using the gray level co-occurrence matrix." NDT & e International 37.7 (2004): 577-588.

[14] Padma, A., and R. Sukanesh. "Automatic classification and segmentation of brain tumor in CT images using optimal dominant gray level run length texture features." International Journal of Advanced Computer Science and Applications 2.10 (2011).

[15] Van Griethuysen, Joost JM, et al. "Computational radiomics system to decode the radiographic phenotype." Cancer research 77.21 (2017): e104-e107.

[16] Hawkins, Douglas M. "The problem of overfitting." Journal of chemical information and computer sciences 44.1 (2004): 1-12.

[17] Hua, Jianping, et al. "Optimal number of features as a function of sample size for various classification rules." Bioinformatics 21.8 (2005): 1509-1515.

[18] American College of Foot and ankle Surgeons. (n.d.). Bone Healing. Retrieved August 04, 2020, from <https://www.foothealthfacts.org/conditions/bone-healing>

[19] Liebsch, Christian, et al. "Patterns of serial rib fractures after blunt chest trauma: An analysis of 380 cases." PloS one 14.12 (2019): e0224105.

[20] Motoyoshi, Isamu, et al. "Image statistics and the perception of surface qualities." Nature 447.7141 (2007): 206-209.

[21] Ahmad, Iftikhar, et al. "Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection." IEEE access 6 (2018): 33789-33795.