Vision system for detection of defects on apples using hyperspectral imaging coupled with neural network and Haar cascade algorithm

P V Balabanov, A G Divin, A S Egorov, V A Yudaev and D A Lyubimova

Department of Mechatronics and Technological Measurements, Tambov State Technical University, 392000, 116 Sovetskaya St, Tambov, Russia

E-mail: egorov.andrey@list.ru

Abstract The article is devoted to the development of algorithms for detecting defective apples transported on a roller conveyor using a vision system. In developing the algorithms, the possibility of classifying various regions of interest (intact and damaged by rot, scab, codling moth, as well as the conveyor) by the principal component method was investigated. When choosing the optimal spectral region for cluster analysis, spectrograms obtained in various spectral ranges, including Vis-NIR (400–1000 nm), NIR (780–1000 nm), and Vis (400–780 nm) were used. The PCA method showed that for the successful classification of the conveyor area, intact, decayed and damaged by the codling moth, it is necessary to use spectrograms in the Vis-NIR range. To classify these ROIs, it was proposed to use a direct distribution neural network with two hidden layers of 128 and 64 layers, respectively, the “relu” activation function in the hidden layers and the “softmax” activation function in the output layer. The optimal network configuration was determined experimentally. This configuration showed a classification accuracy of 0.847 on a test sample of 6,000 apples. Since the samples of spectrograms of scab and stem regions do not differ, for their classification in parallel with the neural network, it was proposed to use the Haar cascade classifier trained on 2000 two-dimensional images of apples in the visible region containing scab and stem regions. The classification accuracy was at least 0.95. The developed algorithm is intended for use in the robotic sorting of apples.

1. Introduction

Vision systems are actively implemented in various fields of human life. In agriculture field, they are used for automated fruits and vegetables quality control, because they provide the possibility of defects quick detection, in particular rot, in the early stages of its appearance [1]. The basis of vision systems is algorithms for control objects images processing. The effectiveness of the algorithms is determined by the percentage of errors in detecting defects and operation speed. The known algorithms are based on machine learning methods [2, 3], for example, support vectors (SVM), nearest neighbors (KNN), and others.

A promising direction is using the hyperspectral method for quality control of fruits and vegetables. It established himself in remote sensing of agricultural crops. Hyperspectral images in a given spectral range with a resolution of no more than 2..3 nm often contain a sufficient information to assess the quality of a test object [4]. Therefore, such methods are used to determine the chemical composition of...
a substance; they make it possible to identify materials, to detect features of an object characterizing its quality.

In the case of fruit quality control, especially in conditions of their transportation on the conveyor, it is necessary to interpret correctly and quickly a large amount of hyperspectral data. Two groups of processing methods for such data can be distinguished: methods involving the operation of binarization (thresholding) and segmentation of hyperspectral images obtained at specific wavelengths [5–7] and methods for obtaining index images and their analysis [8]. In both cases, when obtaining hyperspectral images, cameras with a matrix of sensitive elements or with a line of sensors are used. The last have high speed (over 100 fps), what allows to use them for production purposes, in particular for sorting objects moving on the conveyor.

Algorithms which implement the first group of methods provide for the processing of images obtained in the visible, near-infrared and infrared regions of the spectrum. To detect rot on apples, citrus fruits, the range of 400-1000 Nm is widely used [5]. Within the indicated range, the most contrasting images obtained at given wavelengths are selected (467, 575, 625, 684, 750, 813, 962 nm [6,7]). However, for different types of defects, and even for defects of the same type, different authors give different wavelength ranges at which it is rational to detect them.

The choice of specific wavelengths, in our opinion, is due to the difference in image processing approaches, the difference in the physicochemical properties of the control objects and defects. In addition, as the fruits ripen, their color changes, and, consequently, the intensity of the reflected light changes too. This can be explained by to the fact that the monomeric form of chlorophyll is directly involved in photosynthesis, the amount of which in the fruits during their development decreases by the end of the growing season [11]. Therefore, methods of the second group, involving the use of vegetation indices [12] for spectrograms processing, have been developed. Index images are often used in work with spectral information [13]. Knowing the relationship between the spectral characteristics and the state of the plant tissue of the fruits, makes it possible to use their hyperspectral images to identify their varieties, as well as to control defects, the degree of maturation, withering, etc.

The advantage of vegetation indices using is the easy way of their obtaining and a wide range of tasks which can be solved. In [8], vegetation indices were calculated on the base of apples spectrograms with various injuries (from hailstones, sunburns, scabs), what reflect the content of chlorophyll and the main pigments — anthocyanins and carotenoids. A common chlorophyll index is NDVI, which is based on two independent from other factors sections of the spectral reflection curve [14], whose values are in the range 0.1..0.25 for the surface of apples. In the late stages of apples sunburn, an increase in the anthocyanin index ARI is observed. The nature of the change in the indices in different areas of the apple depends not only on the type of area (defect, peduncle, healthy tissue), but also on the variety of the apple. This is probably due to the difference in surface color and moisture and solid components in the plant tissue. This circumstance makes it difficult to obtain a universal empirical index that allows, by the nature of its change, to make reasonable decisions in the process of sorting apples.

Thus, to identify defects on the apples surface, it seems rational to use the entire spectrogram obtained in the visible and near IR regions (400..1000 nm), rather than its individual sections. Moreover, to analyze a large amount of information in spectrograms, we suggest to use a neural network trained on spectrograms of apples.

2. Materials and methods

2.1 Apple samples preparation

The apples of variety Sinap Orlovsky of 2019 harvest provided by I.V. Michurin Federal Scientific Center were used as samples for the study. Apples of this variety were obtained at the All-Russian Research Institute of Fruit Culture Selection and the All-Union Research Institute of Horticulture named after I.V. Michurina from hybridization in 1955. The average diameter of the used samples was 80 mm. The color of the surface is in range from light green to yellow. The total number of test samples was 300 units, where 100 were intact, the rest contained the following defects: rot, with an area of at least
0.2 sm²; scab spots with an area of more than 0.5 sm²; fruits damaged by the codling moth with at least two damages. Spectrograms of the samples were obtained in the period from 02.10.2020 to 02.15.2020. During this period, the storage temperature of the samples was 20..22 °C, the relative humidity - 60..70%. Upon spectrograms obtaining, the spherical surface of the sample was conventionally divided into segments containing healthy or damaged tissue. The number of segments was chosen arbitrarily and amounted at least 30 for one control object.

2.2 Vision system and data acquisition
The work was performed using the equipment (figure 1) of the “Robotics” collective-use center (registration number 677349) of the Tambov State Technical University.

![Figure 1. Robotic sorting facility](image)

Robotic sorting facility consists of loading hopper 1, where apples are placed and then transported on a roller conveyor through cell 2, where the technical means of the vision system are located. They include Specim FX10 hyperspectral camera 7 and a Basler ace acA1920-155uc visible range camera 6 with a GigE interface, CMOS-matrix Sony IMX174 and a frame rate of 164 frames per second with a resolution of 2.3 MP. The hyperspectral camera has spectral range of 400-1000 nm, spectral band of 224, resolution of 2.5 nm, spatial sampling of 1024 pixels, frequency of 330 frames / sec. The following means were used to illuminate the tested objects: two halogen lamps 8 of the R7SNavigator type with a power of 150 W and a color temperature of 2900 K; Microlight IR Plate-3-850 infrared spotlight 9 with a wavelength of 850 nm; LED lamp 10 Uniel LED / SP / CL ALM01WH, E27, A60, 9W, with a luminous flux of 850 lm. For visualization and processing of spectrograms, a laptop 3 with an Intel Core i5 8300H processor (2.3 GHz - 4.0 GHz in Turbo mode), 8 GB of RAM, and a GeForce GTX 1050 3 GB graphics card were used. The used complex is equipped with an industrial robot UR3 Universal Robots (4 in figure 1), designed to reject defective apples.

2.3 Principal component analysis
The principal component analysis (PCA) was applied to the data recorded in matrix $X$, with $i$ rows determined by the number of spectrogram samples and $j$ columns determined by the spectral range and resolution (in fact, the wavelengths at which the measurements are made). The purpose of the PCA was to test the possibility to classify existing spectrogram samples. The sense of the PCA method is mean that the new variables $t_\alpha$ ($\alpha = 1, ..., A$), called the main components [15] are introduced. They are a linear combination of the initial variables $X_j$ and are defined by the expression $t_\alpha = p_\alpha X_1 + ... + p_\alpha X_j$. Using the introduced variables, the matrix $X$ is represented as the product of two matrices $T$ and $P$ ($X = T \cdot P$), called the matrix of scores and loads, respectively [16]. On the chart of scores, each sample is shown in coordinates $(t_i, t_j)$. Moreover, $i = 1, j = 2$ are most often denoted as PC1 and PC2, respectively. The proximity of points on the graphs means the presence of cross-correlation. Thus, the projections of
spectrogram samples on PC1, PC2 axis with similar features will form clusters. The presence of clusters on the graphs in the coordinates PC1, PC2 allows us to conclude that it is possible to classify spectrogram samples. The graph of loads is used to determine the key wavelengths at which the specific characteristic of the spectrograms are observed (for example, local extrema). In this graph, each $x_i$ variable is displayed as a point in the coordinates $(p_i, p_j)$. To calculate the matrix of scores and loads, we used the Matlab2016 package and the pca function $[\text{loadings, scores}] = \text{pca}(X)$.

2.4 Neural network algorithm
Modern research in the field of determining the properties of control objects by their hyperspectral images shows that, despite the large number of architectures of artificial neural networks [17], fully connected direct distribution neural networks (Deep Feedforward, DFF) are successfully used to solve the problem of classifying an object according to its hyperspectral image. [18-20]. In such networks, the input signal propagates in the forward direction from layer to layer. Multilayer perceptrons include an input layer, one or more hidden layers, an output layer. In this work, we used an input layer consisting of 224 neurons, an output layer of 9 neurons and two inner layers, the optimal number of neurons on each of which was determined empirically. The optimal activation function in hidden layers were selected empirically too.

3. Results and discussion

3.1 Spectra of ROIs of samples
To train the neural network, it was necessary to obtain spectrograms (dependences of the reflection intensity $R$ on the radiation wavelength) for the parts of apple surface damaged by rot, codling moth, scab, as well as undamaged parts of the apple and the conveyor area (for ROI regions of interest). Since the used Specim FX10 camera sensor includes 1,024 sensitive elements arranged in a line, we received 1,024 spectrograms for points lying in the measurement region $[0, x = 300 \text{ mm}]$, the width of which corresponds to the width of the conveyor. Each spectrogram shows the average intensity of reflection from a segment of the measurement region with a width of 2 mm. Therefore, only part of the spectrograms contains the significant information, since it corresponds to the region of interest (ROI).

To determine the informative spectrograms corresponding to the ROI, we analyzed the dependencies $R(x)$ at a wavelength of 750 nm. The dependences $R(x)$ have approximately constant values corresponding to the global minimum for the conveyor region, local minimum in the defect region, and global maximum in the undamaged region of the apple (figure 2). For the neural network training, only spectrograms of areas with the indicated characteristic features were used. It is necessary to note, that local minimum on the $R(x)$ graph cannot exactly indicate the presence of a defect, since the region of the apple containing the peduncle is manifested in a similar way.

![Figure 2. Region of interest (ROI) on the apple and reflectance spectra at 750 nm.](image-url)
A representative sample of the reflection intensity curves for various ROIs in the spectral range 400…1000 nm (Vis-NIR) is shown in figure 3.

![Figure 3. Representative reflectance spectra of ROIs for Vis-NIR spectral region.](image)

As we can be seen from the graphs, the dependences of reflection on wavelength for different regions of the apple are similar. The curves show local minima in the region of 680 nm, corresponding to the radiation absorption by chlorophyll. In this case, the intensity of reflection from normal apple tissue is higher than from damaged or from the area containing the peduncle. The spectrograms of the apple and the conveyor are significantly different, what makes it possible to uniquely identify the apple by the spectrogram.

### 3.2 Principal component clustering analysis based on different spectral regions

In this work, we used the principal component method to determine the possibility of defect classification based on spectrograms obtained for various ROIs. A total of 12,500 spectrogram samples were obtained, including 1,800 conveyor spectrogram samples, 1,700 normal tissue spectrogram samples, and 2,000 rotten tissue spectrogram samples. Each sample was manually assigned a label allowing it to be classified. Figure 4 shows scattering diagrams of samples projections for three classes of spectrograms (conveyor, normal and rotten apple tissue) on the axis PC1, PC2 (first and second main components), respectively. In order to determine the optimal spectral region for cluster analysis, we used spectrograms obtained in various spectral ranges, including Vis-NIR (400–1000 nm), NIR (780–1000 nm), Vis (400–780 nm).

![Figure 4. Principal component clustering analysis of three types of tissues based on two different spectral regions including Vis-NIR(a) and Vis(b), respectively.](image)
As we can see from the figure 4a, the spectrogram samples corresponding to the conveyor class in the Vis-NIR region are well separated from other spectrogram samples. Points which coordinates are projections of spectrogram samples of normal tissue and rotten tissue also form well distinguishable clusters. Figure 4 also shows separate points of the class “normal region” lying in the class “decayed region” and vice versa. This is explained by errors made when labeling samples. The apple region containing the peduncle and sepals is not defective. However, as was noted in Section 3.1, the \( R(x) \) dependence for it has specific local minima, and the spectrograms do not visually differ from the defect or normal tissue region. Therefore, the PCA method was applied, which showed that in the visible range corresponding to the green-orange color (500 ... 630 nm) the spectrogram samples corresponding to the “stem region” class form a separate cluster, distinguishable from “normal region” and “decayed region” (figure 5). Therefore, the spectral regions Vis-NIR and Vis-G were selected for further analysis.

However, the use of all 224 images for analysis in the Vis-NIR spectral region may be impractical. We can find several key wavelengths that affect the classification results. In the figure 6 a graph of loads dependence on wavelength for PC4 is shown. Local extrema on the graph correspond to the wavelengths that contribute greater weight to the result spectrogram. Thus, it was found that the most informative images of the region of interest can be obtained at wavelengths of 566 nm, 639 nm, 679 nm, 709 nm, 962 nm.

The reflection intensities at the indicated wavelengths can be used to classify defects. However, defects in apple tissue caused by scab could not be classified by PCA. Points which coordinates are samples projections of spectrograms for tissues containing these defects do not form a separate cluster,
but are located in the “stem region” region. With this in mind, we concluded that it is unreasonable to include samples of scab spectrograms in the neural network training set and we used the Haar cascade to classify the area containing the peduncle and sepals.

3.3. Classification of stem region by Haar cascade

To detect the area of the apple containing the peduncle and sepals, Viola-Jones algorithms were used [21]. The search of the control object is carried out on an image with a size of 1920x1080 pixels received from a Basler camera by scanning it with a 24 × 24 px window. To provide the smallest recognition error, the specified size was scaled with a coefficient K in the range from 5 to 15. A part of the image with window sizes was passed through a cascade of 12 layers.

The Haar features, representing the spectral coefficients of the Haar orthogonal decomposition and defined as differences of pixel intensities sums in neighboring areas of the image, were calculated on each layer. The cascade represents 12 layers, on each of which there is a strong classifier, providing for the combination of many Haar features (weak classifiers). Moreover, the number of weak classifiers on each subsequent layer increases. On each of the classifier layers, the correspondence of the analyzed part of the image to the object of search. If the detection result is negative (equal to 0) on one of the layers, then the detection stops and the scanning window moves to the next area. If the result is equal to 1, then the image inside the scanning window will be considered as an object. To optimize the method productivity, the Adaboost algorithm [22], which represents a sequence of actions for training weak classifiers, choosing the most accurate and compiling their linear combination, was used. To train the cascade, 1000 images of the apples region containing the peduncle or sepals (positive images), as well as 1000 images of the apple surface that did not contain these objects or defects (negative images) were used. To train the cascade, the standard opencv_traincascade application was used. To run the application on the Windows operating system, the following command is used:

```
opencv_traincascade -data Apple -vec pos.vec -bg neg.txt -numPos 1000 -numNeg 1000 -numStages 12 –precalcValBufSize 4096 -precalcIdxBufSize 4096 -featureType HAAR -w 24 -h 24m 24.
```

The data parameter sets the name of the folder where after the training completion the trained cascade will be saved. The list of paths to negative images is specified by the –bg parameter. The number of positive and negative images is determined by numPos and numNeg parameters, respectively. The number of cascade layers is specified by numStages parameter. The amount of memory allocated for data processing is set by precalcValBufSize and precalcIdxBufSize parameters. The featureType parameter determines the type of weak classifiers.

To increase the recognition accuracy of the sought areas of the apple, we determined the optimal values of the parameter K in the range 5 .. 15 and minNeighbors from the sequence 1, 5, 10, 15, 20, corresponding to the minimum percentage of recognition errors (figure 7). Errors were considered as cases when the peduncle or sepals were not found in the apple image. Cases of false detection of the or sepals if they are not in the image were also considered errors.

![Figure 7. Stem region recognition errors.](image-url)
3.4. Classification results for all samples by neural network

To classify the control objects, a direct distribution neural network with two hidden layers is used. When choosing the network configuration, the experimental studies on the dependence of the classification accuracy of control objects on the network configuration parameters were conducted. The configuration parameters were following: the number of neurons in hidden layers, the type of activation function in the network layers, the number of epochs during training, and the sample size during network training. The following values were checked: the number of neurons 64, 80, 96, 112, 128 for the first layer, and 16, 32, 48, 64 for the second layer. Number of epochs: 10, 20, 30. Batch size: 250, 500, 750, 1000. The preliminary studies have shown that in order to achieve the classification accuracy of at least 50%, the following activation functions must be used: “sigmoid”, “softmax”, “relu”. For all layers, combinations of these functions were checked. The total number of verified network configurations was 6480.

The neural network was trained on 1000 hyperspectral images of the conveyor area and 11000 hyperspectral images of apples, including healthy samples, defects (rot, damage by the moth), as well as areas of apple tissue containing the sepals and peduncle. The data were divided into training and test samples in the ratio of 80% and 20%, respectively. For the software implementation of the neural network, the Python 3.7 programming language and the Keras library were used.

As a result of experimental studies (a full report on all configurations is available at https://github.com/egorovandrey/hyperspectral_AI_results) and calculations, a direct distribution neural network with two hidden layers, the activation function “relu” in hidden layers and the activation function “softmax” in the output layer was chosen. The number of neurons in the hidden layers was 128 and 64 for the first and second layer, respectively. The number of training epochs was 30. The batch size is 250. This configuration showed the classification accuracy of 0.847 on the test sample. A further increase in the number of neurons or epochs in the selected network configuration did not allow to increase the classification accuracy, while the training and classification process performed with lower rate. The final configuration of the neural network is shown in the figure 8.

![Figure 8. Neural network architecture.](image)

3.5. Apples defects detecting algorithm for robotic sorting system

The main component of the vision system software is the neural network (module ROIs neural network classification on figure 9), which allows to determine the object class in the camera’s field of view using spectrograms obtained from the Specim camera. Classes 0, 1, 2 at the output of the module correspond to the regions of interest conveyor, normal region and stem region, respectively. Classes 3 and 5 correspond to defects on the apple surface (rot, damage by the moth). The ROI containing the defect of the scab type and the Stem region is determined by the neural network as class 2. In this case, the decision on whether the ROI belongs to the class Stem region (2) or Scab region (4) is made by a True / False signal from the output of the logical AND block (&) according to the following algorithm. If the True signal is at the output of the Stem region detection module, then the ROI belongs to the class 2, otherwise to class the 4. The reject command and coordinates of the rejected object are sent to the manipulator only if the ROI of object belongs to classes 3, 4, 5.
4. Conclusions

A recognition algorithm for apples transported on the conveyor was developed and allows to determine fruits coordinates and classify them depending on the types of surface defects. The classes of defects are defined - decayed, scab, wormy regions, as well as the classes that are not related to defects - normal, stem, conveyor regions. For the mentioned ROIs, 12,500 one-dimensional hyperspectral images (spectrograms) were obtained in the range of 400..1000 nm with a resolution of 2.3 nm. To analyze the spectrograms, the principal component method was applied and showed the possibility of classifying regions of interest. In the Vis-NIR range, spectrogram samples corresponding to the conveyor, normal, and decayed classes form separate clusters. In the Vis-G range, samples of the normal, stem, and decayed classes also form separate clusters. These features led to the conclusion that it is necessary to take into account the entire Vis-NIR range for the classification of ROIs. The exception was samples of spectrograms of the scab and stem classes, which do not differ significantly from each other. Therefore, the Haar cascade classifier, trained on 2000 two-dimensional apples images in the visible region containing scab and stem regions, was used to classify these ROIs. The classification accuracy was at least 0.95 at \( \text{minNeighbors} = 1..3 \) and \( K = 7..10 \). For the classification of stem, normal, decayed, wormy and conveyor regions, we used the direct distribution neural network with two hidden layers of 128 and 64 neurons, the “relu” activation function in the hidden layers and the “softmax” activation function in the output layer. The highest classification accuracy of 0.85 was obtained on a test sample of 6000 fruits with a number of training epochs 30 and Batch size 250.

The work was performed using the equipment of the "Robotics" collective-use center (registration number 677349) of the Tambov State Technical University.

This work was supported by the Ministry of Science and Higher Education of the Russian Federation under the Agreement No. 05.604.21.0240, unique identifier RFMEFI60419X0240.

References

[1] Buslov E Yu, Zon B A and Kornienko A V Optical method of sorting marketable grain according to gluten content 2010 Journal of optical technology 77(6) 391-3
[2] Bhargava A and Bansal A Automatic Detection and Grading of Multiple Fruits by Machine
Learning 2019 Food Analytical Methods 13(3) 751-61
[3] Liakos K G, Busato P, Moshou D. Machine Learning in Agriculture: A Review 2018 SENSORS 18(8) 138-46
[4] Chen Y R, Chao K L and Kim M S Machine vision technology for agricultural applications 2000 International Conference on Engineering and Technological Sciences 36(2-3) 173-91
[5] Li J B, Luo W, Wang Z and Fan S. Early detection of decay on apples using hyperspectral reflectance imaging combining both principal component analysis and improved watershed segmentation method 2019 Postharvest Biology and Technology 149 235-46
[6] Huang W Q, Li J B, Wang Q Y and Chen L P Development of a multispectral imaging system for online detection of bruises on apples 2015 Food Eng. 145 62-71
[7] Baranowski P, Mazurek W, Wozniak J and Majewska U. Detection of early bruises in apples using Hyperspectral data and thermal imaging 2012 Food Eng. 110(3) 345-55
[8] Merzlyak M N, Solovchenko A E and Gitelson A A Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit 2003 Postharvest Biology and Technology 27(2) 197-211
[9] Patankar A B and Tayade P A. Application of Computer Vision in Agriculture 2015 International Journal of Electronics and Computer Science Engineering 4(3) 238-44
[10] Yuzhen L, Yuping H and Renfu L. Innovative Hyperspectral Imaging-Based Techniques for Quality Evaluation of Fruits and Vegetables 2017 Applied Science 7 189
[11] Tuccio L, Agati G and Grassini G. Non-destructive fluorescence sensing for applications in precision viticulture 2015 International Conference on Biophotonics 79-82
[12] Mousaei Sanjerehei M Assessment of spectral vegetation indices for estimating vegetation cover in arid and semi-arid shrublands 2014 Range Management and Agroforestry 35 (1) 91-100
[13] Cohen W B, Maiersperger T K, Gower S T and Turner D P An improved strategy for regression of biophysical variables and Landsat ETM+ data 2003 Remote Sensing of Environment 84 (4) 561-71
[14] Crippen R E Calculating the vegetation index faster 1990 Remote Sensing of Environment 34 71-3
[15] Roweis S EM Algorithms for PCA and SPCA 1998 In Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems 10 626-32
[16] Jolliffe I T 2002 Principal component analysis. 2nd ed. (Springer)
[17] He M and Huang R. 2005 Feature selection for hyperspectral data classification using double parallel feedforward neural networks. Lecture notes in computer science pp 58-66
[18] Fernandez-Redondo M, Hernandez-Espinosa C and Torres-Sospedra J. 2004 Hyperspectral image classification by ensembles of multilayer feedforward networks (IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541))
[19] Marpu P R, Gamba P and Niemeyer I 2009 Hyperspectral data classification using an ensemble of class-dependent neural networks (First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing)
[20] Viola P and Jones M J Robust real-time face detection 2004 International Journal of Computer Vision 57 137-54
[21] Rohith Gandhi. 2018 Boosting Algorithms: AdaBoost, Gradient Boosting and XGBoost. Retrieved from hackernoon.com