**BIO542: Machine Learning for Biomedical Applications (MLBA)**

**(8th December 2024, End-sem exam)**

**Maximum Marks: 60 Duration: 90 Minutes**

**Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Admission No: \_\_\_\_\_\_\_\_\_\_\_\_\_**

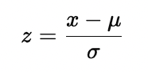
**Instructions:** This question paper have two sections, A and B. Attempt any 14 questions from section A, each question carries 2 marks (Total 28 marks). Attempt any 8 questions from section B, each question carries 4 marks (Total 32 marks). Write all answers in answer sheet only.

**Section A**

1. **Name the tool used for genotype imputation and haplotyping using WGS sequence.**

Ans: MaCH

1. **Write formula for standardization**

Ans: 

Where:

* z is the standardized value.
* x is the original value.
* μ is the mean of the dataset.
* σ is the standard deviation of the dataset.

1. **Which feature selection method uses Lasso regularization (Filter, Wrapper, Embedded)**

Ans: Embedded method

1. **In which year, *Caenorhabditis elegans* has been sequenced and what is its genome size.**

Ans: 1998, 97 Mb or 19,000 genes

1. **Write full form of TCGA and its year of initiation.**

Ans: The Cancer Genome Atlas and 2006

1. **How many drugs and cell lines (information) have been in maintained in CancerDR.**

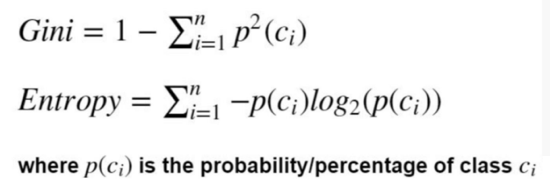
Ans: 148 drugs and 952 cell lines

1. **Write name of any two boosting techniques.**

Ans: AdaBoost, Gradient Boosting, XGBoost, CatBoost, LightGBM

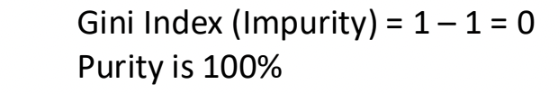
1. **Write the formula of gini index and entropy?**

Ans:



1. **A bag contains 5 blue balls, what will be entropy and gini index.**

Ans:

****



1. **Write full form of deep learning techniques LSTM and RBM**

Ans: Long Short-Term Memory networks, Restricted Boltzmann Machines

1. **Describe importance of pooling in CNN.**

Ans: Pooling reduces the complexity. We can subsample the image to make it smaller and with fewer parameters that can characterize the image.

1. **Write name of four deep learning frameworks**

Ans: TensorFlow, Keras, Pytorch, Theano, DL4J, Caffe, Chainer, Microsoft CNTK

1. **Write applications of software AC2DGel.**

Ans: Analysis and comparison of 2D gels

1. **Write the full form of any two: SIFT, RIFT, HoG, GLOH.**

Ans: SIFT: Scale-invariant feature transform, RIFT: Rotation-invariant feature transform, HoG: Hough transform, GLOH: Gradient Location and Orientation Histogram

1. **Sort following techniques based on their interpretability (ascending order): RF, ANN, LR, SVM**

Ans: ANN, RF, SVM, LR

1. **What is the full form of LIME and SHAP**

Ans: Local Interpretable Model-agnostic Explanations, (SHapley Additive exPlanations)

1. **List two ML based models come under category of opaque models**

Ans: RF, SVM

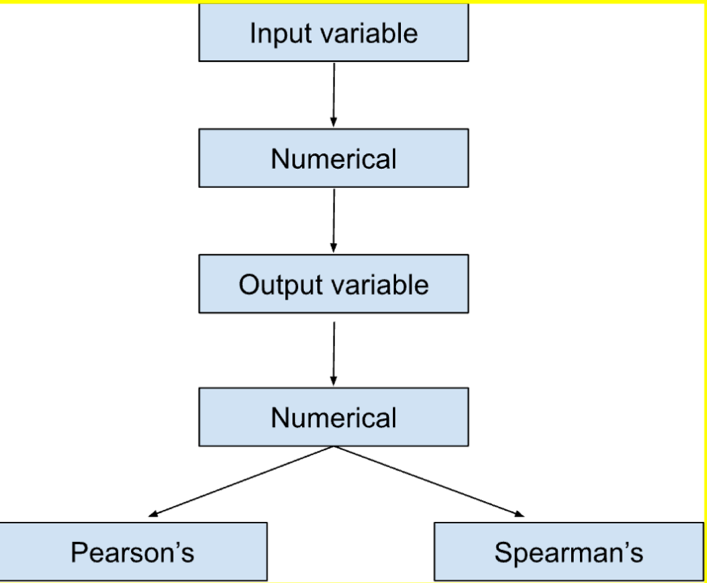
1. **Give two examples of rule-based models.**

Ans: Decision Tree, Random forest, CART: Classification and Regression Trees)

**Section-B**

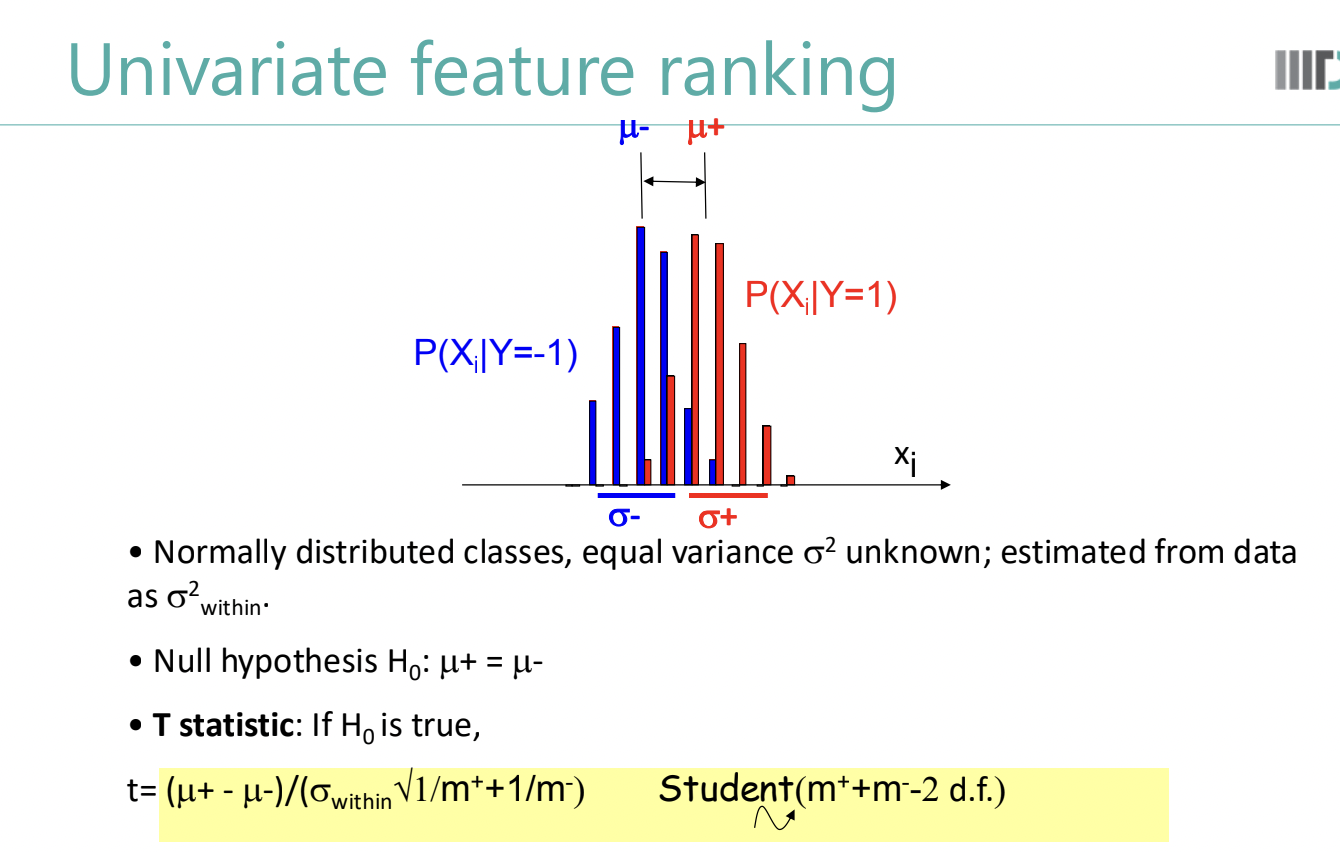
1. **Show univariate feature selection by flowchart (Input and Output variables are numerical).**

Ans:

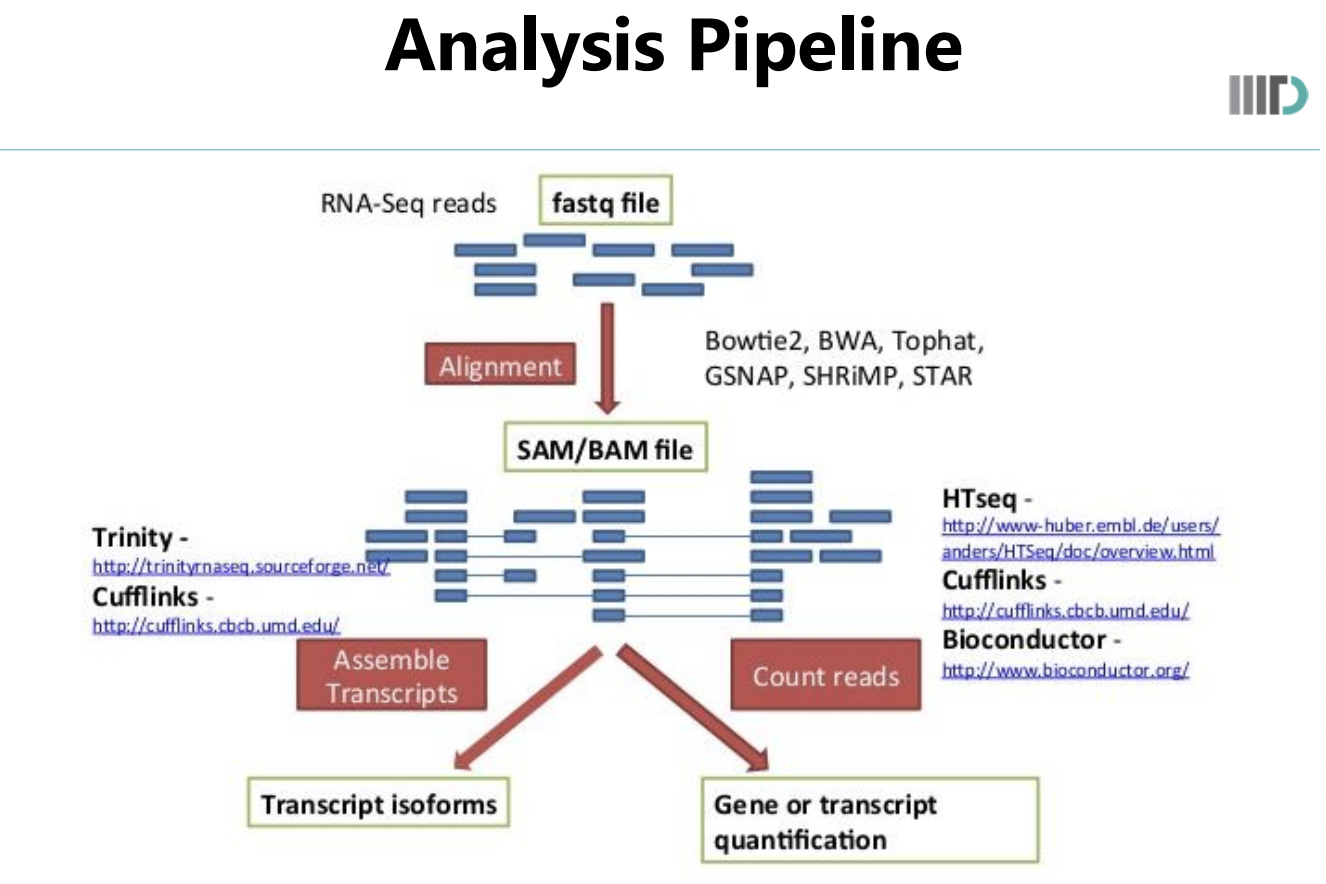


1. **Describe univariate feature ranking with graph and t-statistics**

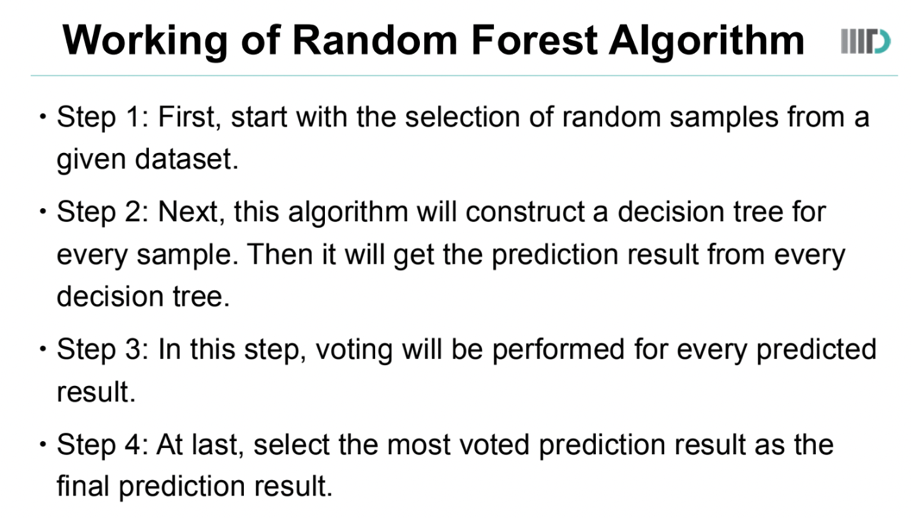
Ans: Univariate feature ranking is a technique used to evaluate the significance of each feature individually with respect to the target variable. It involves calculating a specific statistical measure for each feature and then ranking them based on their scores. Here it is asked to show it with the t-statistics.



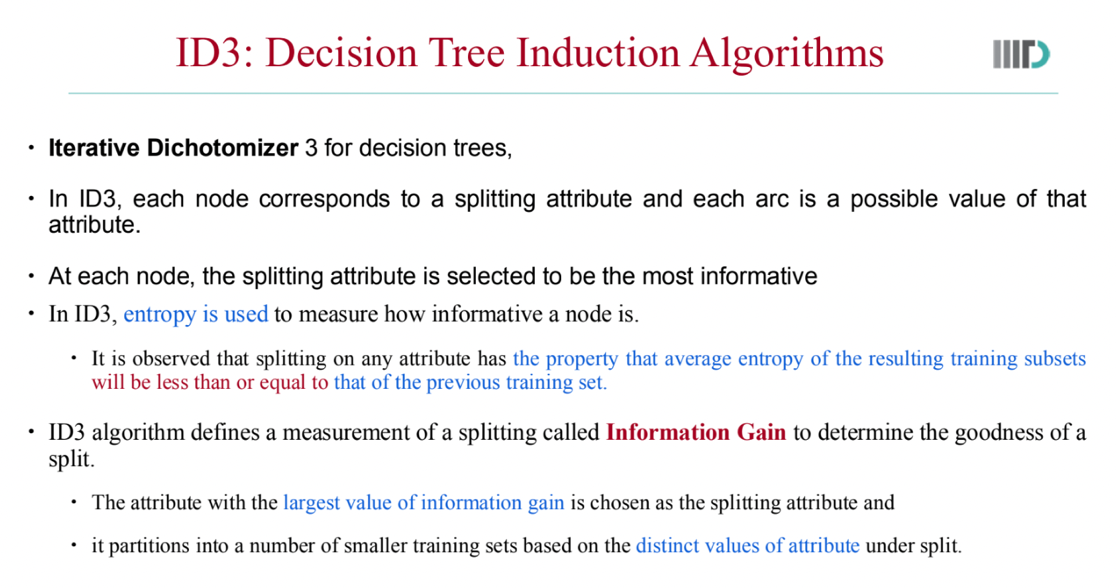
1. **Draw schematically the Analysis Pipeline of RNA-Seq.**

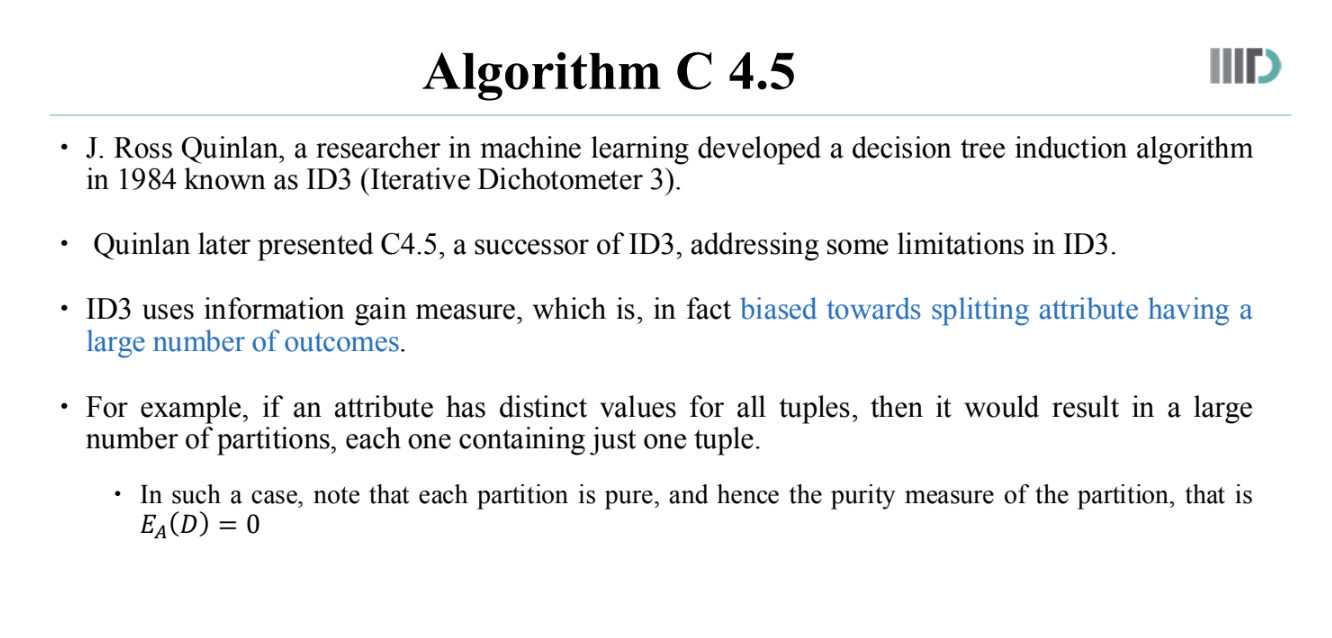


1. **Write four major steps involved in Random Forest Algorithm.**

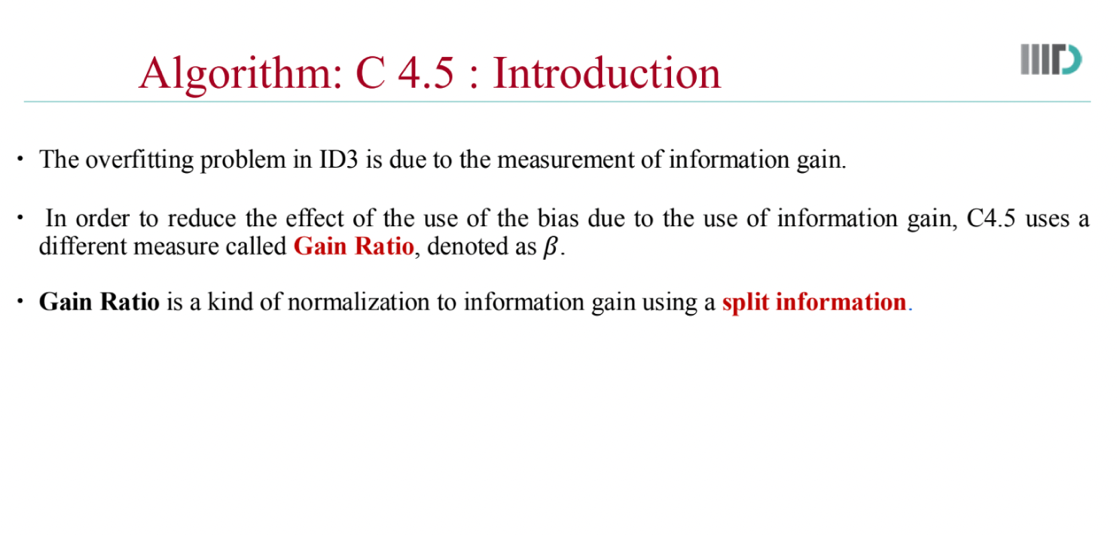


1. **Briefly describe and compare ID3 and C 4.5 algorithm**

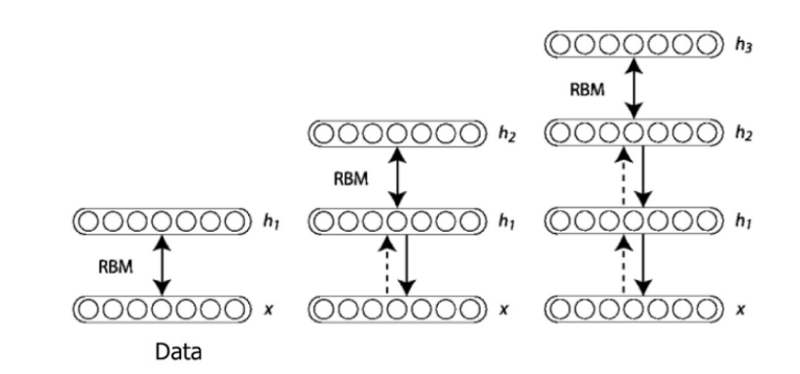




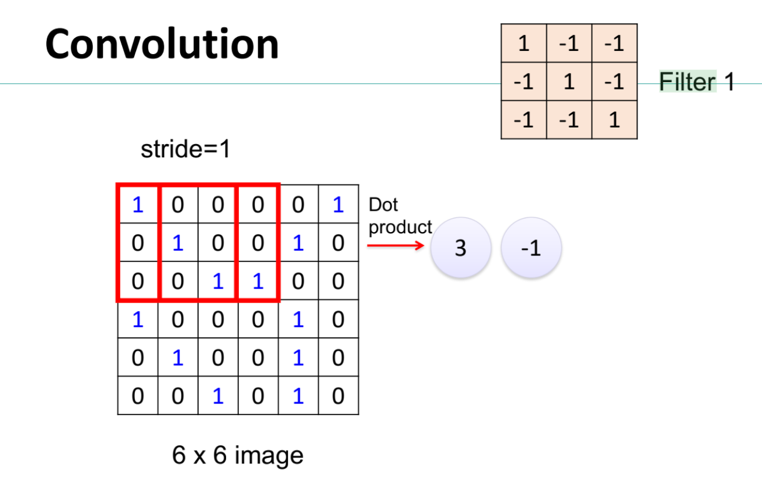
Comparison:



1. **Show DBN architecture by flowchart, show RBM and hidden layers**

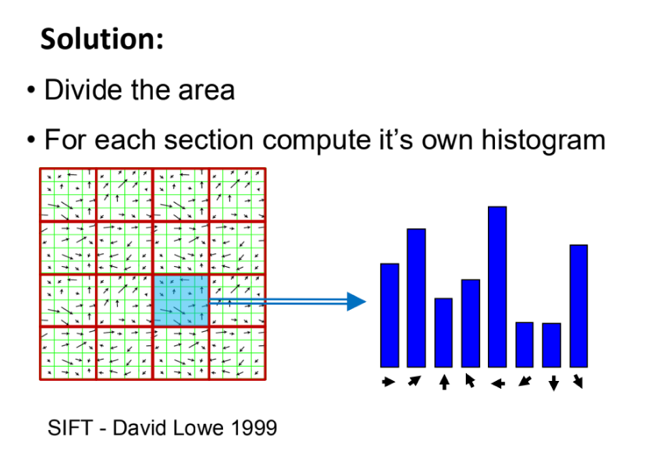


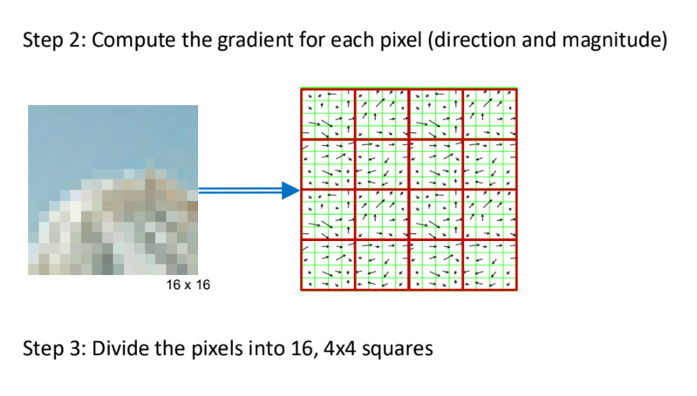
1. **In CNN, show multiplication of an image matric with a filter**

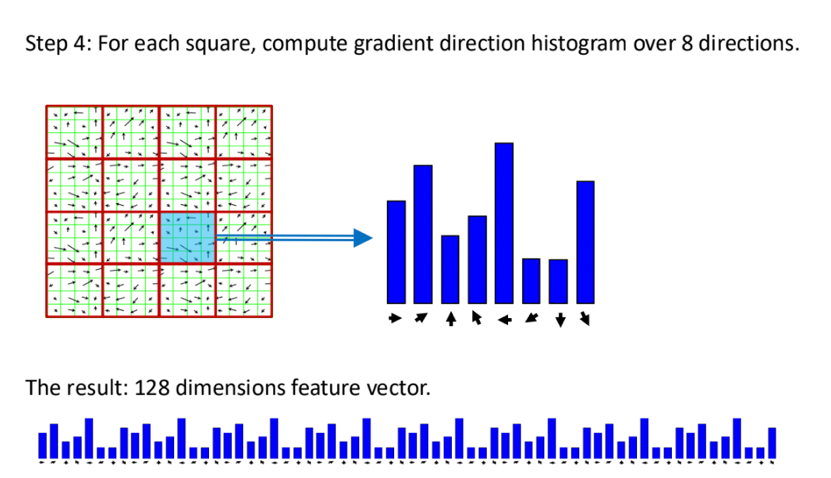




1. **Explain how SIFT descriptors are computed with steps and vector size.**









1. **Write code for computing SIFT descriptors using OpenCV.**



1. **Show taxonomy of interpretable machine learning methods by flowchart**

