REVIEW 1

ARAVIND PRASAD P B

M Tech CSE 2019-2020

120033002

September 23, 2019

ABOUT THE BASE PAPER

- Milos Ilic, Sinisa Ilic, Srdjan Jovic, Stefan Panic,
- Early cherry fruit pathogen disease detection based on data mining prediction",2018,Computers and Electronics in Agriculture(Elseiver).
- Impact Factor **2.427**

AGENDA

- Introduction
- Problem Statement
- System Model
- Data Visualization
- Program for Image Prepossessing
- Program for Edge Detector
- Images of Edge Detector
- Result Discussion
- Conclusion
- Future Work
- References

INTRODUCTION

- First goal of this paper is **Data prepossessing, Visualization, Outlier detection** and Reduction of redundancies among instances.
- Second goal is to create prediction models based on refined data after Prepossessing by using different **data mining prediction techniques**.
- Third goal is comparison of **prediction models** and **selection of the best model**.
- Determined by the basis of **statistical errors** and **data mining errors** for the each model
- Focuses on two data sets such as,First data set type is containing data used for **prediction model training**
- Second data set type represents data that are used for **Prediction model testing** and **Accuracy evaluation**

MOTIVATIONS

- Different Mathematical techniques for **Data Processing** and **Data Protection**
- Important diseases in cherry Fruit is **Monilinia Laxa** and **Coccomyces Hiemalis**
- Achieved Prediction accuracy is **95.8%**

PROBLEM STATEMENT

- Prediction can be performed based on weather data and data that represent **specific plant or disease** characteristic that are maintained
- Weather conditions are favorable for disease development which can greatly reduce the problems and helps in **fruit protection**
- By the **least expensive** for collection are weather parameters and each disease can **infect fruit tree** if specific weather conditions are satisfied
- Weather factors are selected in **every month** as prediction parameters.

SYSTEM MODEL

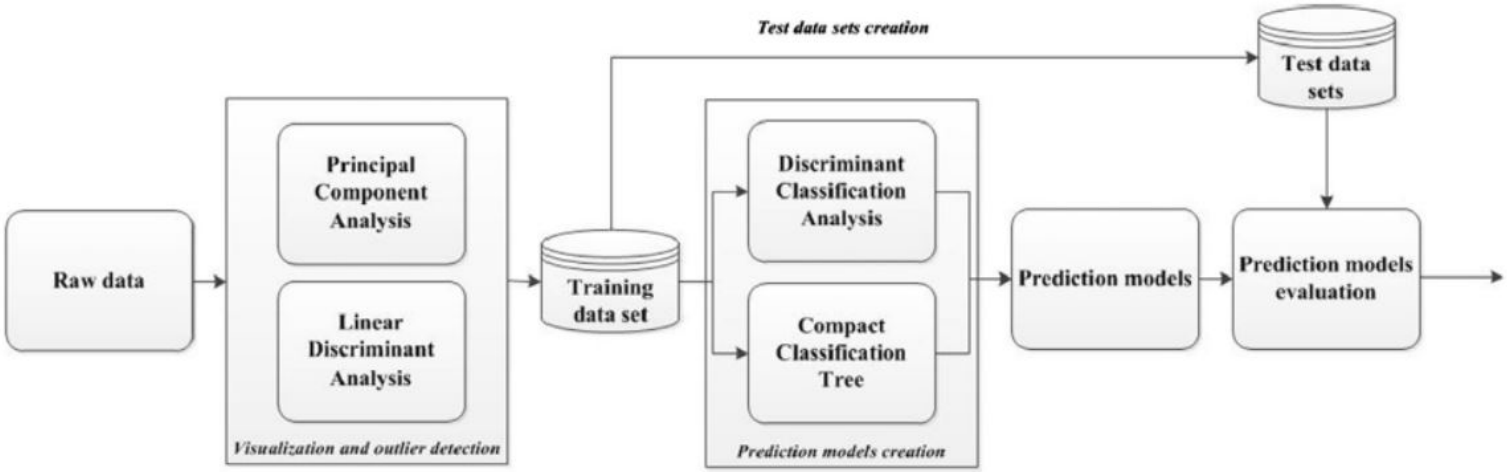

Figure: System Model for Mutual relations for implemented Method

Source: Computers and Electronics in Agriculture (Elseiver)-2018

- Data prepossessing
- Evaluation prediction model creation
- Prediction of future class attribute values

MODEL TYPES AND EXPLANATION

- File System used here is **single .xlsx file** ,which provides input to the System Visualization and Outlier Detection module
- Two groups of methods **PCA** and **Discriminant analysis**.
- Some **outliers** are detected using PCA and remaining using **Discriminant analysis**.
- Each of them has own advantages and flaws

CLASSIFICATION ANALYSIS

- Varies types of analysis are present in both **Classification and Clustering**
- Deals with tree Structure (i.e) Classification tree,Decision tree
- Developed predictive models are mainly evaluated classification processing.
- Accuracy of prediction is evaluated by **Comparing the actual value of the class attribute** and the value of predicted class attributes in the test data set.

PRINCIPAL COMPONENT ANALYSIS

- It mainly used abundantly in varies analysis such as **Neuroscience to Computer graphics** because it is a simple, non-parametric method of extracting relevant information from raw data sets.
- Additional effort PCA provides a roadmap to **reduce a complex data set** to a lower dimension to reveal the sometimes **hidden**, **simplified structure format**.
- Significance of the **DPCA matrix** is having high dimensional data and reduce it to just a few principal components hopefully.
- Capture variations of the data and allow **Visualization,Patterns** and **Structures**

- Discriminant analysis consist of a **instances of observations** with known **group membership** together with their continuous variables
- Prior probability can have one of the three choices: **Uniform**,**Empirical** and **Numeric vector**.
- Uniform prior probability of class k is 1 over the **total number of classes** present.
- Empirical prior probability of class k is the number of training instances of **class k** divided by the **total number of training instances**.
- Numeric vector prior probability of class k is the **jth element of the prior vector**.

COMPACT CLASSIFICATION TREE

- Classification tree splits nodes either impurity or nodeerror,but only three of them are widely used **Gini splitting rule, Deviance** and **Two-ing splitting rule**
- Decision trees are then used to **classify new data**. They are used for each instance from learning data set the class value is known.
- It may be provided by user or calculated in accordance with **some exogenous rule**
- Common Notations are used as,**tp** be a parent node and **tl**,**tr** for left and right child nodes of parent node(tp).
- Variable matrix **X** with **M** number of variables xj and **N** observations made in it.

EVALUATION PARAMETERS

Variable name	Min	Max	Mean	Variance	StdDev
Min temp	-2	28	12.97	24.52	4.95
Max temp	5	26.26	47.14	6.87	40
Mean temp	3	33	19.60	29.93	5.47
Average humidity	0.33	0.97	0.68	0.02	0.13
Rainfall	0	30	1.05	10.56	3.25
Wind speed	3	27	7.59	16.63	4.08

Tabulation:Evaluation Parameter

DATA VISUALIZATION

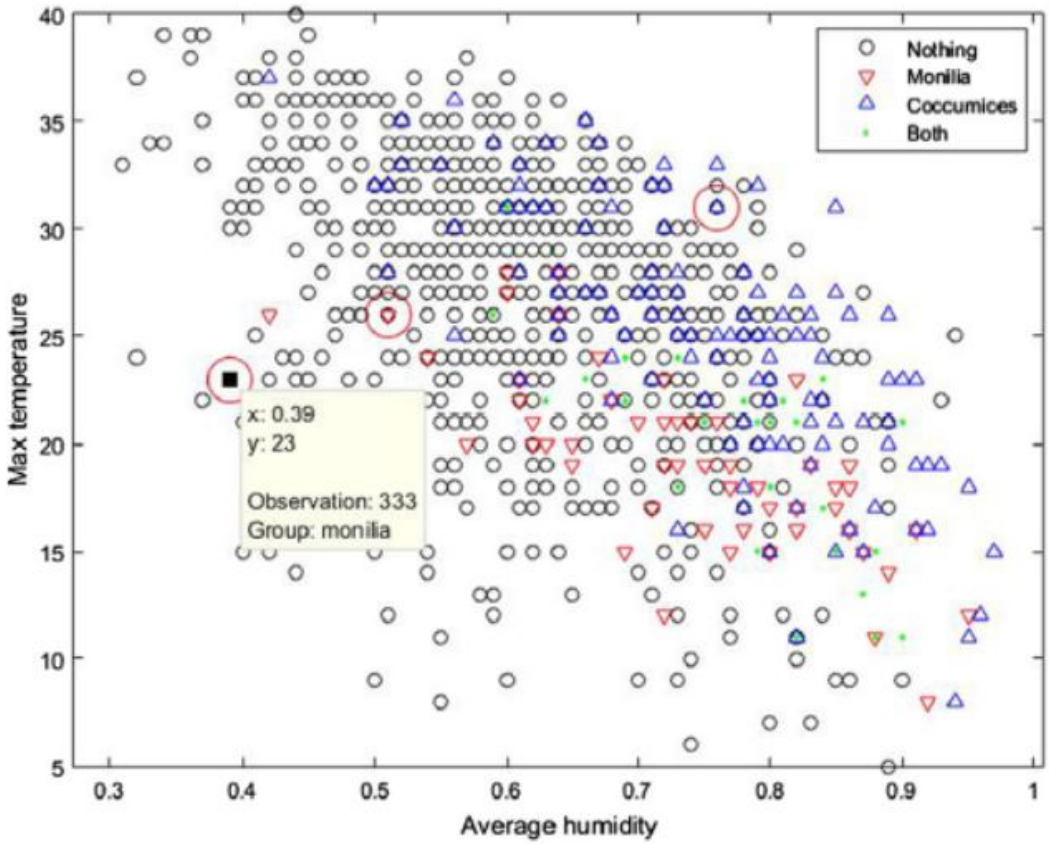

Figure: 2D Visualization of Training Data Set

Source: Computers and Electronics in Agriculture(Elseiver)-2018

3 D DATA VISUALIZATION

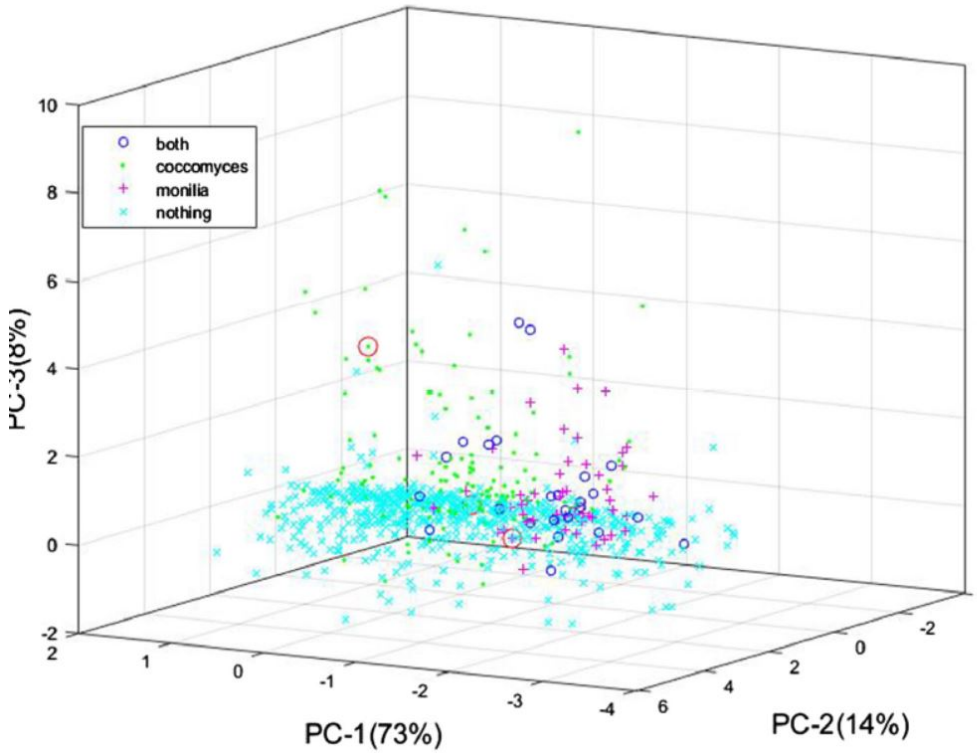

Figure: 3D Visualization

Source: Computers and Electronics in Agriculture(Elseiver)-2018

Program for Image Preprocessing

```
from PIL import Image
import PIL
from PIL import Image

image=Image.open("C:\\Users\\Aravind Prasad\\Desktop\\MRICE\\Bacterial leaf blight\\a.jpg")
print(image.format)
print(image.mode)
print(image.size)
image.show()
image.thumbnail((100,100))
print(image.size)
cropped=image.crop((100,10
cropped.show()
cropped.show()
cropped.show()
```

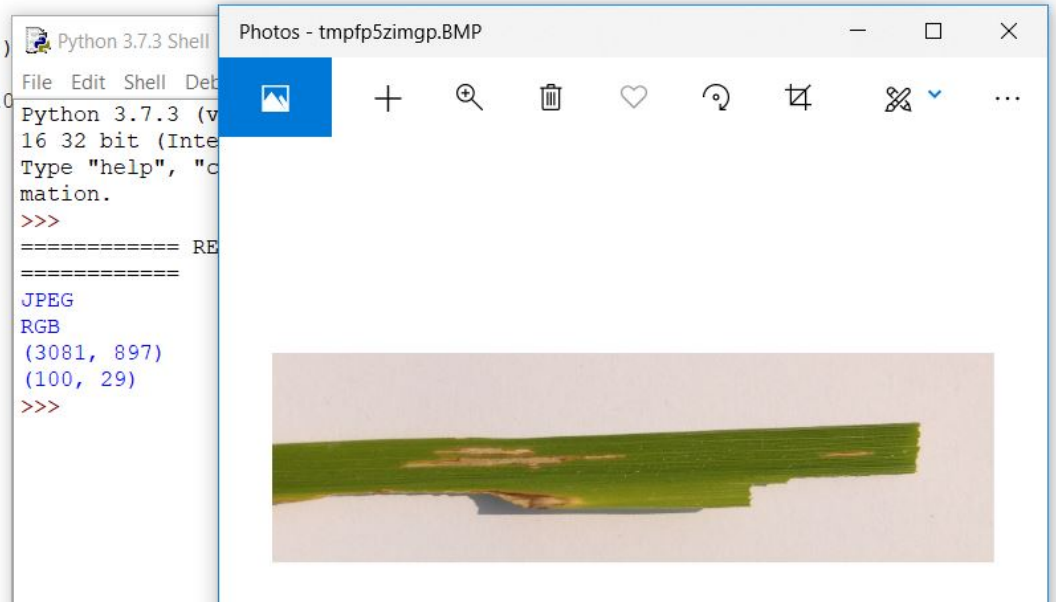

Figure: Image Prepossessing

Program for Edge Detector

```
import cv2
from matplotlib import pyplot as plt
img = cv2.imread('C:\\Users\\Aravind Prasad\\Desktop\\MRICE\\Bacterial leaf blight\\d.JPG')
grey = cv2.cvtColor(img,cv2.COLOR_BGR2GRAY)
blur = cv2.GaussianBlur(grey, (5, 5), 0)
edge = cv2.Canny(blur,100,200)
dil = cv2.dilate(edge, None)
plt.subplot(231),plt.imshow(img,cmap = 'gray')
plt.title('Original Image'), plt.xticks([]), plt.yticks([])
plt.subplot(232),plt.imshow(grey,cmap = 'gray')
plt.title('Grey Image'), plt.xticks([]), plt.yticks([])
plt.subplot(233),plt.imshow(blur,cmap = 'gray')
plt.title('Blur Image'), plt.xticks([]), plt.yticks([])
plt.subplot(234),plt.imshow(edge,cmap = 'gray')
plt.title('Edge Image'), plt.xticks([]), plt.yticks([])
plt.subplot(235),plt.imshow(dil,cmap = 'gray')
plt.title('Dilate Image'), plt.xticks([]), plt.yticks([])
plt.show()
```

Figure: Program for Edge Detector

Images of Edge Detector

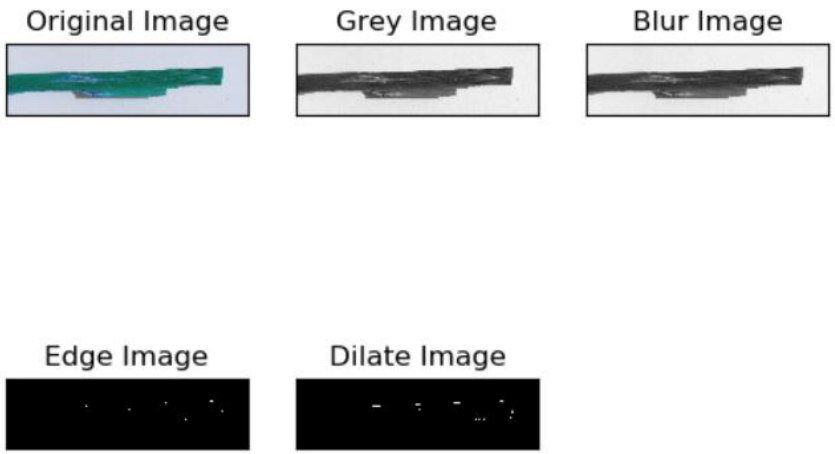

Figure: Edge Detector

RESULT DISCUSSIONS

- Re substitution error is error rate obtained as **0.0286** and for 10-fold cross validation is **0.0817**.
- Trees that classifies the original training set well, but the structure of the tree is sensitive and its performance on new data is likely to **degrade** for **test data set**.
- Easier for **finding simple tree** for unknown data set performs better than a complex one.

CONCLUSIONS

- Predicting time for infection occurrences, the chemical protection is **more efficient** so that **farmers could save money** and on the other side more importantly,it **provides healthier food** due to reduced chemical treatments.
- Implemented algorithms are used for **infection prediction** on the two specific cherry fruit diseases.
- Evaluation of all implemented algorithms was performed under the **same conditions** and the **same evaluation parameters** were observed.
- Main parameter for the development of the disease is **minimal rainfall** (i.e) if there is no rain, necessary conditions for the development of the disease are not fulfilled.

FUTURE WORK

- Authors future plan is to extend the tool with **automated weather data collection** through the network of meteorological stations.
- Another upgrade in the future will be **automatic farmer notification of possible infection** through the mobile or Internet network.

REFERENCES

1 Arsevska, E., Roche, M., Hendrikx, P., Chavernac, D.,Falala, S., Lancelot, R., Dufour, B.,2016. Identification of terms for detecting early signals of emerging infectious disease outbreaks on the web. Comput. Electron. Agric. 123, 104115.

2 Lucky, M., Christina, C., Kevin, G., Peter, O.S., 2016. Predicting pre-planting risk of Stagonospora nodorum blotch in winter wheat using machine learning models. Frontiers Plant Sci. 7.

3 Babu, M.S.P., Srinivasa, B.R., 2010. Leaves recognition using back-propagation neural network - advice for pest and disease control on crops. Technical report, Department of Computer Science Systems Engineering, Andhra University, India.

REFERENCES (CONT..)

4 Chakraborty, S., Ghosh, R., Ghosh, M., Fernandes, C.D., Charchar, M.J., Kelemu, S., 2004.Weather-based prediction of anthracnose severity using artificial neural network models. Plant. Pathol. 53, 375386. http://dx.doi.org/10.1111/j.1365-3059.2004.01044.

5 Roman, T., 2004. Classification and Regression Trees (CART) Theory and Applications, A Master Thesis, Center of Applied Statistics and Economics Humboldt University, Berlin, pp. 140.

6 Toghroli, A., Mohammadhassani, M., Suhatril, M., Shariati, M., Ibrahim, Z., 2014.Prediction of shear capacity of channel shear connectors using the ANFIS model. Steel Compos. Struct. 17 (5), 623639.

THANK YOU