

BREAST CANCER DISANOSIS USING MACHINE LEARNING METHOD

Guanqing Li, Jiawen Zhang, Zhiyuan Lou

University of California San Diego, La Jolla, CA 92093-0238,

ABSTRACT

Breast cancer is one of the most common cancer and is causing a huge number of deaths in women. The high incidence and mortality of breast cancer is due to its considerably low accuracy of diagnosis. In this paper, we explore machine learning models that can be applied to help increasing the accuracy of the diagnosis of breast cancer. The main problem of the project is to detect breast cancer based on a set of features calculated from a digitized image of the Fine Needle Aspiration (FNA) of a breast mass from a patient. We present a diagnosis model using both traditional and deep learning machine learning models. Classic machine learning models including Logistic Regression, Nearest Neighbor, Support Vector Machine, etc. are tested on the Breast Cancer Wisconsin dataset. Additionally, we applied feature selection and neural network model to improve the performance of the system. This paper demonstrates that machine learning models can be used for an automatic diagnosis for breast cancer.

Index Terms—Breast Cancer Diagnosis, Logistic Regression, Nearest Neighbor, Support Vector Machines, Kernel Support Vector Machines, Naive Bayes, Decision Tree Algorithm, Random Forest Algorithm

1. INTRODUCTION

Breast cancer is one of the most common cancer in women and the second leading cause of women's cancer death[1]. Despite the lack of effective treatment, the low accuracy of diagnosis is also a major cause of the high incidence and mortality of breast cancer. Mammography is a traditional method used for diagnosing breast cancer. According to UCHealth's report, only 78% of breast cancer can be accurately diagnosed by mammography [2]. Many cases such as doctors' negligence or incompetence in addition to a mammography error may also result in a late diagnosis or misdiagnosis, which can be considered a cause of breast cancer death [3]. In the long term, early-stage diagnosis could significantly increase the survival rate of breast cancer [4], therefore, it is important to improve the accuracy of breast cancer diagnosis.

Machine learning has been applied in medical diagnosis in a large number of papers [5]. In order to increase the accuracy of breast cancer diagnosis, we aim to use machine learning models and choose the model with higher performance.

Breast Cancer Wisconsin is a widely used dataset provided by UC Irvine machine learning repository. In this paper, we will train our models using this dataset.

The input of our algorithm is a set of features calculated from a digitized image of the Fine Needle Aspiration (FNA) of a breast mass from a patient. We will then use seven traditional methods including Logistic Regression, Nearest Neighbor (k-NN), Support Vector Machines (SVM), Kernel Support Vector Machines (KSVM), Naive Bayes, Decision Tree and Random Forest Algorithm, and a deep learning method to predict whether the case is benign or malignant.

2. RELATED WORK

There have been many studies applying different machine learning techniques on medical analysis. In terms of traditional machine learning methods, Chaurasia et al.[6] used Simple Logistic to reduce the dimension of feature space and applied RepTree and RBF Network to evaluate the performance. Dubey et al. used K-means algorithm to evaluate the impact of clustering using centroid initialization and achieved 92% average positive prediction accuracy [7]. Classification and regression trees (CART) classifier with feature selection and bagging technique was implemented to predict breast cancer in [8]. Wang et al. [9] compared four classifiers: Naive Bayes, Decision Tree, Support Vector Machine and k-nearest neighbor for classification of cancer using gene expression data. These traditional machine learning models has the advantage of low design complexity, but it is not capable of dealing with complex data. In the project, we are going to apply models including these previous mentioned algorithms on the Breast Cancer Wisconsin dataset and compare the performances.

Many deep learning models have also been developed in this objective. In [10], problems in multiple datasets were discussed and Partial Likelihood Artificial Neural Network is applied for prediction of cancer survival. Purwar et al. [11] proposed a hybrid model using a combination of K-means clustering with Multilayer Perceptron with promising results for various medical dataset. Another new classification algorithm for detection of breast abnormalities in digital mammograms using Particle Swarm Optimized Wavelet Neural Network (PSOWNN) is investigated in [12]. These deep learning models are capable of modeling complex and high dimen-

sional data. However, the computational complexity is higher and training time may be long. In this paper we will build a neural network to apply to the same Breast Cancer Wisconsin dataset and compare the performance with traditional models.

3. DATASET AND FEATURES

The dataset we used is Breast Cancer Wisconsin dataset which is a widely used dataset in study. It contains 699 instances with 9 features and 2 classes (benign and malignant.) The class distribution is as follows: 458 benign (65.5%) and 241 malignant (34.5%). The features are: 1) radius (mean of distances from center to points on the perimeter). 2) texture (standard deviation of gray-scale values). 3) perimeter. 4) area. 5) smoothness (local variation in radius lengths). 6) compactness ($perimeter^2/area - 1.0$). 7) concavity (severity of concave portions of the contour). 8) concave points (number of concave portions of the contour). 9) symmetry. 10) fractal dimension ("coastline approximation" - 1).

3.1. Data Preprocessing

We first cleaned the dataset by removing samples with empty values. There are 683 samples after removing invalid samples. We then realized that the dataset with 683 samples is rather small. To enhance the dataset, we generated a new dataset by copying original dataset and add Gaussian noise to it. Afterward, we appended the generated dataset to the original dataset. This process doubles the dataset to 1398 samples. Then we rescaled data to $[0,1]$ using MinMaxScaler.

By plotting scatter matrix of the features, we realized the data is significantly right skewed, which may make the model biased towards the majority of the features. Thus, we took the square root of the feature data to mitigate the data skew problem. The result is shown in Fig 1.

4. METHODS

We used 7 traditional models for the classification of breast cancer cases. Feature selection is applied to increase the rate of accurate prediction. Additionally, deep learning model is also built for the diagnosis system. Finally, we compare the performance of all the models applied and choose the one with the highest performance.

4.1. Traditional models

(1) Logistic Regression (LR): Logistic regression predicts the probability of the default class (e.g. Class 2 in this case) and transforms the probability into a binary value (0 or 1) for classification using "sigmoid" function as shown in Equation 1.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

(2) K-Nearest Neighbor (k-NN): K-nearest neighbor assigns a case to the class that is most common among its k nearest neighbors. The distance between the case and its neighbor is measured by using distance functions like Euclidean: $D_{Euclidean} = \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$, Manhattan: $D_{Manhattan} = \sum_{i=1}^k |x_i - y_i|$ and Minkowski: $D_{Minkowski} = (\sum_{i=1}^k (|x_i - y_i|^q))^{\frac{1}{q}}$

(3) Support Vector Machine (SVM): Support Vector Machine finds an optimal hyperplane that best separates the classes based on the support vectors. The function of kernel for SVM is to take data as input and transform it into the required form.[13] The kernel function used in SVM model is linear function as shown in Eq.2

$$k(x_i, x_j) = a < x_i, x_j > + b \quad (2)$$

(4) Kernel Support Vector Machines (Kernel SVM): The kernel SVM in this paper is the SVM algorithm that uses Gaussian radial basis function (RBF) as kernel. The RBF is shown in Eq.3.

$$k(x_i, x_j) = exp(-\gamma ||x_i - x_j||^2) \quad (3)$$

(5) Naive Bayes (NB): NB algorithm makes classifications using the Maximum A Posteriori decision rule (Eq.4) in a Bayesian setting.

$$y = argmax_{c_i} (P(c_i) \prod_{j=1}^n P(x_j | c_i)) \quad (4)$$

where c_i is one of the classes and x_j is one of the features.

(6) Decision Tree (DT): The decision tree are presented with a tree structure. The test objects are classified by their feature values. A node in a decision tree represents an instance, outcomes of the test represented by branch, and the leaf node epitomized the class label[14].

(7) Random Forest (RF): Random forest is a set of individual decision trees. Each decision tree spits out a class prediction. It decides the class of the test object by aggregating the votes from different decision trees[15].

4.2. Feature selection

In many machine learning algorithms, there is a decrease of accuracy when the number of features is redundant [16]. In order to improve the accuracy of the models and avoid overfitting, we performed feature selection on the data. For the traditional models, we used two techniques to select features from the dataset. For the Decision Tree and Random Forest model, we generate the feature importance of the last training result and choose features accordingly. Given the importance of the j th feature I_j , we drop two features that has minimum feature importance: $\hat{j}_{drop} = argmin_j (I_j)$.

For the remaining five models, we used the correlation matrix with heatmap visualization. Correlation represents

how the features in the dataset are related to each other. By using the heatmap visualization, it is easier to identify which features are highly correlated. Using the seaborn library, we could plot the heatmap for better view. For each group of highly correlated features, we choose only one feature to represent all that are in the group. This way most of the information in the features is reserved, and the redundant information is dropped to avoid overfitting.

4.3. Deep learning model

The structure is shown in Fig. 2. The first hidden layer has 9 neurons, followed by GaussianNoise layer to improve robustness, and dropout layer to reduce overfitting. Then the process is repeated and the last dense layer has 1 neuron. The activation function is relu.

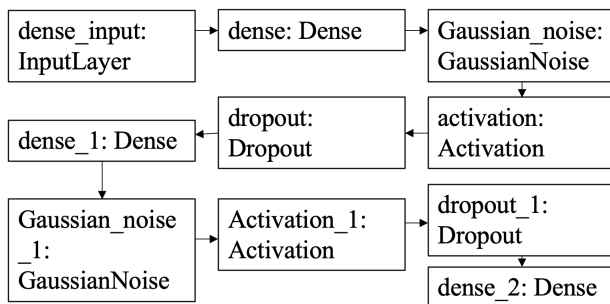


Fig. 1: Deep learning model

5. EXPERIMENTS AND DISCUSSION

5.1. Experiment setting

The server we used is Google Colab. We used Scikit-learn machine learning package for the implementation and evaluation of traditional models as well as preprocessing of data. TensorFlow Core v2.2.0 is used for developing the neural network model. For the visualization of results, we used Seaborn and Matplotlib visualization. We applied our models on the enhanced Breast Cancer Wisconsin dataset with a 60%-40% training-testing split (838/560). For traditional methods, additional 70%-30% and 80%-20% training-testing splits were applied. For k-nearest neighbor model, the number of neighbors was set to be 5 and the "Minkowski" distance function was used. For SVM model, linear kernel was used. For kernel SVM, the kernel function was set to be RBF. For decision tree model, the function to measure the quality of a split was set to be entropy. For the random forest model, the number of estimators was set to be 10 and the function to measure the quality of a split was set to be entropy. All these parameters were set as above based on [17]. With these parameters, the seven traditional models yield good results. The performance of the models are evaluated by performance

metrics including accuracy, F score and confusion matrices. These metrics exhibit similar result, hence we will mainly talk about accuracy as the measure of model performance.

5.2. Results

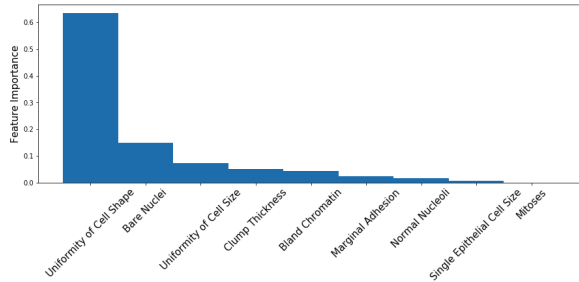
| Model | Accuracy (60/40) | Accuracy (70/30) | Accuracy (80/20) | Confusion Matrix (60/40) | | |
|---------------------|------------------|------------------|------------------|--------------------------|--------|-----------|
| Logistic Regression | 95.53% | 95% | 93.57% | Benign | 368 | 11 |
| | | | | Malignant | 14 | 167 |
| | | | | | Benign | Malignant |
| K-nearest Neighbor | 94.82% | 94.76% | 95% | Benign | 367 | 12 |
| | | | | Malignant | 17 | 164 |
| | | | | | Benign | Malignant |
| SVM | 95.71% | 95.48% | 93.57% | Benign | 366 | 13 |
| | | | | Malignant | 11 | 170 |
| | | | | | Benign | Malignant |
| Kernel SVM | 95.89% | 96.19% | 95.36% | Benign | 365 | 14 |
| | | | | Malignant | 9 | 172 |
| | | | | | Benign | Malignant |
| Naïve Bayes | 95.54% | 95.71% | 95.71% | Benign | 362 | 17 |
| | | | | Malignant | 8 | 173 |
| | | | | | Benign | Malignant |
| Decision Tree | 95.54% | 94.76% | 96.07% | Benign | 370 | 9 |
| | | | | Malignant | 16 | 165 |
| | | | | | Benign | Malignant |
| Random Forest | 96.61% | 96.19% | 96.43% | Benign | 366 | 13 |
| | | | | Malignant | 6 | 175 |
| | | | | | Benign | Malignant |

Fig. 2: Traditional models' accuracy and confusion matrix

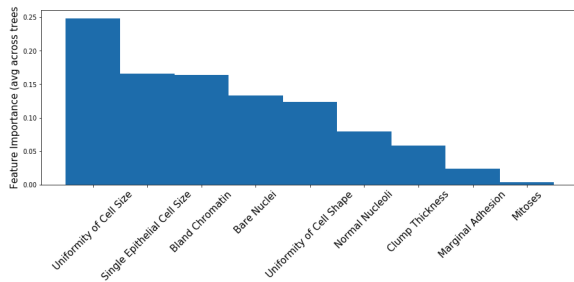
Fig.2 shows the accuracies of the seven traditional models with three different training-test splits (60%-40%, 70%-30% and 80%-20%). The confusion matrix of each model with a 60%-40% training-test split is also shown in Fig.2. As a result, all the seven models achieved a very high accuracy around 95%. With all three different training-test splits, the RF model achieved the highest accuracy of over 96%. As shown in the confusion matrix of random forest, the model predicted accurately on malignant cases. Only 6 cases were misclassified. For logistic regression and SVM, the accuracy decreases with the increasing percentage of the training set. LR model did a relatively good work in classifying benign cases, but the number of misclassified malignant cases is relatively large. SVM model did a bit better than LR in classifying malignant cases, but did a bit worse in classifying benign cases. The change of training set ratio did not affect much on k-NN, Kernel SVM and NB models. NB and Kernel SVM models achieved a high accuracy in classifying malignant cases while k-NN did not. But the misclassify rates of benign cases of the two models were high while k-NN model's was relatively low. For DT model, the change of training set ratio had a slight effect on the accuracy. The highest accuracy of 96% was achieved with 80% training set, and the lowest accuracy of 94% was achieved with 60% training set. DT model has the second highest misclassify rate of malignant case, but it achieved the highest accuracy in classifying benign cases among all 7 models.

The feature importance of Decision Tree and Random Forest model are shown in Fig.3. We chose to drop the two features with minimum feature importance and trained the

model again with remaining features. The accuracy of the model before and after feature selection is shown in Table 1. The accuracy of the DT model is increase by applying feature selection while the accuracy of RF model is decrease. From the feature importance map we can see the importance of the last two features of the DT model are significantly lower than the RF algorithm. The RF model distributes the importance over features more evenly, which makes dropping features have negative impact on the model accuracy.



(a) Decision Tree model



(b) Random Forest model

Fig. 3: Feature importances

| | Before FS | After FS |
|---------------|-----------|----------|
| Decision Tree | 0.9660 | 0.9696 |
| Random Forest | 0.9678 | 0.9625 |

Table 1: Accuracy of models before and after feature selection by feature importance (FS: feature selection)

| | Before FS | After FS |
|-------------------|-----------|----------|
| Linear Regression | 0.9571 | 0.9589 |
| k-NN | 0.9500 | 0.9553 |

Table 2: Accuracy of models before and after feature selection by correlation matrix (FS: feature selection)

The heat map visualization is shown in Fig. 5. The features that have high correlations will have a lighter color in the corresponding grid. We replaced group of features that have high correlation with only one from the group. Table 2 shows that applying feature selection have increased the accuracy of both the LR and the k-NN models.

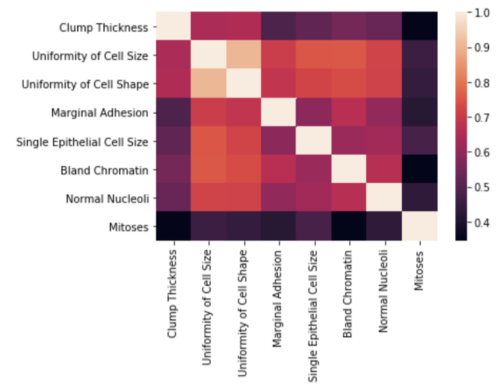


Fig. 4: Heat map visualization of correlation matrix

| Train/Validation data split | 60/40 | 70/30 | 80/20 | Confusion matrix (60/40) | Benign | Malignant |
|-----------------------------|--------|--------|--------|--------------------------|--------|-----------|
| Accuracy | 96.96% | 97.14% | 97.50% | Benign | 357 | 10 |
| | | | | Malignant | 7 | 186 |

Fig. 5: Deep learning model’s accuracy and confusion matrix

For the deep learning method, it has an accuracy of 96.96% under 60%-40% data split, which is slightly higher than traditional method. The accuracy increases with the number of training data. From the confusion matrix, we can see that the deep learning model missed 17 cases in all, as shown in Fig 5.

6. CONCLUSION AND FUTURE WORK

In conclusion, comparing with the 78% accuracy of mam-mography, all the seven traditional models achieved a higher accuracy of around 95% for breast cancer diagnosis. Among all traditional models, the random forest model provided the best result (more than 96% accuracy) to our dataset. In order to improve the performance, we experiment with feature selection using correlation matrix and feature importance. It results in better accuracies when removing certain features for certain model. We discussed the drawback of this dataset being relatively small and containing skewed data, proposed a way to mitigate the issue by appending the same samples with additional Gaussian random noise. A neural network is built to apply to the new, enhanced dataset to achieve an accuracy of 96.96% under 60%-40% split. Future work includes tuning the hyperparameters of the current model as well as testing other deep learning method/architectures to increase model accuracy. Overall, we presented machine learning models that can be applied in breast cancer diagnosis to improve the accuracy and therefore assist early diagnosis of breast cancer.

7. REFERENCES

- [1] A. C. Society, "How common is breast cancer?." <https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html>, 2011.
- [2] UCHealth, "How accurate are mammograms?." <https://www.uchealth.org/today/how-accurate-are-mammograms/>, 2015.
- [3] "Breast cancer misdiagnosis and mammography errors." <https://hackmd.io/@shaochia/Hk6fwgI1r?type=view1-Introduction>, 2011.
- [4] L. Wang, "Early diagnosis of breast cancer," *Sensors*, vol. 17, no. 7, p. 1572, 2017.
- [5] K. R. Foster, R. Koprowski, and J. D. Skufca, "Machine learning, medical diagnosis, and biomedical engineering research-commentary," *Biomedical engineering online*, vol. 13, no. 1, p. 94, 2014.
- [6] V. Chaurasia and S. Pal, "Data mining techniques: to predict and resolve breast cancer survivability," *International Journal of Computer Science and Mobile Computing IJCSMC*, vol. 3, no. 1, pp. 10–22, 2014.
- [7] A. K. Dubey, U. Gupta, and S. Jain, "Analysis of k-means clustering approach on the breast cancer wisconsin dataset," *International journal of computer assisted radiology and surgery*, vol. 11, no. 11, pp. 2033–2047, 2016.
- [8] D. Lavanya and K. U. Rani, "Ensemble decision tree classifier for breast cancer data," *International Journal of Information Technology Convergence and Services*, vol. 2, no. 1, p. 17, 2012.
- [9] X. Wang and O. Gotoh, "A robust gene selection method for microarray-based cancer classification," *Cancer informatics*, vol. 9, pp. CIN–S3794, 2010.
- [10] R. Marshall, "Artificial neural networks in cancer management,"
- [11] A. Purwar and S. K. Singh, "Hybrid prediction model with missing value imputation for medical data," *Expert Systems with Applications*, vol. 42, no. 13, pp. 5621–5631, 2015.
- [12] J. Dheebea, N. A. Singh, and S. T. Selvi, "Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach," *Journal of biomedical informatics*, vol. 49, pp. 45–52, 2014.
- [13] D. Team, "Kernel functions-introduction to svm kernel examples." <https://data-flair.training/blogs/svm-kernel-functions/>, 2018.
- [14] S. B. Ranjit Panigrahi, "Classification and analysis of facebook metrics dataset using supervised classifiers," *Social Network Analytics*, 2019.
- [15] A. Chakure, "Random forest classification and its implementation in python." <https://towardsdatascience.com/random-forest-classification-and-its-implementation-d5d840d0bead0>, 2019.
- [16] M. B. Kursu, W. R. Rudnicki, *et al.*, "Feature selection with the boruta package," *J Stat Softw*, vol. 36, no. 11, pp. 1–13, 2010.
- [17] vishabh goel, "Building a simple machine learning model on breast cancer data." <https://towardsdatascience.com/building-a-simple-machine-learning-model-on-breast-cancer-data-eca4b3b99fa3>, 2018.

Contributions

Guanqing Li

She worked on part of training with traditional model, feature selection, programming, literature review and report writing.

Jiawen Zhang

He worked on part of training with traditional model, generation of performance metrics, programming, literature review and report writing.

Zhiyuan Lou

He worked on implementation of the deep learning model, dataset augmentation, programming, literature review and report writing.

Reply to Reviews

Critiques by Group 9

Clear and concise explanation about the dataset.

Good choice to go through the data and discard data points that were either invalid, with respect to some quantitative or qualitative metric, or incomplete. This is a critical step in data preprocessing, since you want your model to be trained on data that represents the type of data you expect to be fed as input.

Good choice to also unskew the data. Machine learning models must generalize well, and by checking the distribution of the training data, and modifying it to remove any skewness, you ensure that your model is trained on a well-balanced dataset.

Our response: Thanks!

It doesn't seem any feature extraction was done to extract new features. I would be interested to see what features you can extract from the dataset and see if they can help improve your models, or if they are detrimental. Even if they are detrimental, it is a result that is helpful.

Our response: The features from the dataset are derived from the result of FNA, which is extracted from image. Also as you can see from the result, the features we have now are adequate for a high accuracy. I think another round of feature extraction will be redundant. We have also shown feature selection where reducing the number of features improve the accuracy nby a little.

It is interesting that the Random Forest model was able to achieve 100% accuracy with the 80%/20% training-validation split. It is unclear if this was training accuracy, validation accuracy, or testing accuracy though I suspect it is training accuracy. Were hyperparameters tuned to achieve this result, or were the default hyperparameters used?

Our response: It is validation accuracy.

Good choice to add both the dropout layers to reduce overfitting, and adding the Gaussian noise layer to improve the robustness of your model. Did you also consider using the leaky rectified linear unit instead of just the rectified linear unit? I've seen the leaky unit used more because it helps reduce vanishing gradient problems.

Our response: Leaky rectified linear unit has significantly better performance over sigmoid regarding to the vanishing gradient problems. However, we did not find significant difference between leaky rectified linear unit and standard rectified linear unit.

The choice of models and deep learning architecture is simple enough that it can be applied for use in real-life medical devices. Though a lot of machine learning focuses on methods for improving classification rates, many applications of machine learning are either too complex or expensive to be

implemented in real-world devices. This is either due to hardware or economical constraints. By implementing a model that is not complex, but performs very well, this makes it usable. To that end, I think this group did a great job of choosing models that have high accuracy, but are not too complex.

Our response: Thanks!

The accuracies for all models were incredibly high. This is great, but in general most machine learning models are unable to perform with this high of an accuracy unless the training dataset contains already curated, carefully filtered data. This seems to be the case here, but this is not a criticism of the group's efforts.

Our response: We have talked about how the problem of dataset being too small is solved in the data augmentation section

This is more a criticism of the dataset itself, and not necessarily the work done by the group. The dataset used for this project consisted of features that were most likely manually (or semi-manually) labeled by a medical professional. This would assist with breast cancer diagnosis when the features used by the model are collectable. Fine-needle aspiration is a minimally invasive procedure and can be done in a very short amount of time (no more than 10-15 minutes). Compare this to MRIs, while not invasive, take an extensive amount of time and can clog up vital resources. Comparing the two methods, fine-needle aspiration is the more practical approach, but I am curious to see if using a convolutional neural network, where the MRI image is fed as input, can help increase accuracy. This plays into the tradeoff between accuracy vs. model complexity.

Our response: Thanks! We will consider trying to take MRI images as inputs in the future.

The project is very interesting and solves a problem that is critical to the survival of cancer patients: early detection. It is well-known and accepted by many medical professionals that the earlier cancer is detected, the better the chances of survival. Additionally, the project shows that

classification of breast cancer can be done using data collected through minimally invasive means. This is a stark improvement from methods used in the past, like a biopsy where the patient would have to undergo surgery just to collect the cancerous cells for study (though in some cases, surgery is the only option, like with brain cancer).

Our response: Thanks!

Critiques by group 17

This talk thoroughly explained the dataset and gave a detailed comparison of all the traditional methods. They also use a deep learning method to solve this problem. Using feature selection is also a great idea to improve the performance. The experiment is very sufficient.

Our response: Thanks!

Some improvements/unclear: – Why the Linear Regression model performed better in most of the case, why others not? Maybe lack some explanations about the reasons why these models behave differently in this dataset.

Our response: Sorry. We made a mistake in the presentation. It's logistic regression model, not linear regression model. Because the original dataset is small, the difference of accuracy may caused by only one misclassified case. So, we cannot really say LR model works better. Actually the performance of all seven model are very similar. So we enlarged the dataset in the final report, we found that RF worked better in all cases. RF works better in our project

- Why the Random Forest Algorithm got a better result when the size of training set increased?

Our response: This is just a coincidence. After we enlarged our dataset, the increasing size of training set did not result in a better result of the RF model. The accuracy of RF is stable at around 96%.

-Similar, please explain why deep learning method is not a best choice in this case.

Our response: The accuracy of machine learning methods using same training/validation split are quite similar, the small difference may caused by the small dataset issue. Also, in the presentation, the deep learning method uses augmented dataset with random Gaussian noise, while the other methods use original dataset.

Critiques by group 50

The project aims to improve the prediction of breast cancer diagnosis using machine learning. The literature survey indicates machine learning classification methods such as Support Vector Machines, k-nearest neighbors, Naïve-Bayesian classification techniques has been previously applied for this application. The problem presented here is a classical binary classification problem to predict a benign or malignant tumor for breast cancer diagnosis.

First, the project applies seven classification algorithms and compares the respective accuracies by experimenting with the test train split 70-30 vs 80-20. The RF algorithm surprisingly has a 100% accuracy in the latter configuration while the others algorithm do not trail far behind, with accuracies ranging from 94-99%. Second, they experiment with feature selection to train the RF model, which results in better accuracies when removing certain unimportant features. They lay out certain drawback of this dataset which is small in size (699 samples) and contains some skewed data as well. To alleviate these issues, they proceed to double the sample size by appending the same dataset to the original, except with additional Gaussian random noise added to the appended data and scale the features. Then, a deep learning method is applied to the new, larger dataset to achieve an accuracy of 96.7%. Future work includes tuning the hyper-parameters of the current model as well as testing other deep

learning method/architectures to increase model accuracy. Overall, the project rationale and methodology were clearly explained.

Our response: Thanks!

Some questions include- -Does the deep learning method work with the original dataset of 699 samples? If so, what is the accuracy?

Our response: Yes, the accuracy is 96.43 percent%

-Has the appending of the same dataset to the original albeit with gaussian noise been used in previous studies? It seems rather convenient to do so, but maybe justifiable only under certain circumstances. However, the deep learning model (and possibly other methods) after feature engineering will probably be able to filter noise out completely, in which case you might be artificially inflating the accuracy of your model? Just a thought.

Our response: We can prevent this by only using original dataset in test, and the dataset with Gaussian noise is only used for training.

- How do the other machine learning methods performing with the larger (1300 sample) dataset? Your RF model is already 100% accurate, so how is that affected?

Our response: See Section Result for the performance.