Face Mask Detection

Project Members: Shania Dhani, Xuejin Gao, Michelle Lucero

Project Description (Q3**):** Design an algorithm that most accurately determines whether an individual is wearing a mask, not wearing a mask, or wearing a mask incorrectly. Given the suddenness of the current COVID-19 pandemic, relatively few research papers that address face mask coverage are available. It should be noted however, that the machine learning models currently available, while minimal, yield promising results.

Currently, the paper, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic," does not account for wearing masks incorrectly (Loey, M et al., 2021). We plan to address this third classification in our research and use our findings to build our own face-detection-model that accounts for this category. To accomplish this we plan to use a combination of different datasets available to create a new dataset that is labelled for mask v. no mask v. incorrect wearing of a mask. We will also add in a few more classifiers not used in the original study such as KNN and different types of CNN to gauge which models are the best predictors. We hope to use these insights to create our own face-detection model.

Member's Role and Contribution (Q4**):**

Shania will be in charge of determining how to fit the Mask/No Mask/Improper Wear of Mask dataset using a KNN Classifier and a Decision Trees Model. The goal will be to find the most optimal settings for these models based on our accuracy and error analysis metrics (ex: best 'k' for KNN). The models that are most comparable to our baseline study's accuracy of 99.64% will be analyzed and considered in the construction of our own final model (Loey, M et al., 2021). It should be noted that the study mentioned ran their model against three different datasets, achieving 100% accuracy over one of the three.

Xuejin will be learning and reading more on the different types of CNN(CNN, R-CNN, Fast R-CNN) and applying these models to the Face Mask Dataset. The objective is to find a model that has a low computational time with really good accuracy. After determining the best algorithm for CNN's, discuss with teammates on how to optimize to get better performances.

Michelle will be collecting and cleaning the image dataset, adding facial landmark features, and implementing/analyzing SVM and Deep Learning models.

Member	**Role**	**Contribution (Expected)**	**CSCI 353 Related Topics, Questions/Solution Approaches**
Shania D.	KNN and Decision Tree Model	- Test different 'k' and 'distance' algorithms on KNN Classifier to determine what aspects result in	- KNN Classifier, sFold, Decision Trees, Classification

	Implementation & Analyzation; Recorder/Paper Contributor	the best performance metrics without overfitting. - Analyze characteristics that produce an accuracy comparable to our foundational study and extract them for our own face-mask detection model.. - Research Decision Trees, and repeat process above.	(What class does each image belong to), Supervised Learning, Binary V. Multi Classification
Xuejin G.	Responsible for learning about CNN and implementing it into our project.	- Apply different types of CNN and measure their performances	- Neural Networks, CNN, Naive Bayes, comparing performance measures
Michelle L.	Data Preprocessing, Implementing/Analyzing SVM and Deep Learning	- Collecting/cleaning mask/no mask dataset - Creating features/target columns - Apply SVM - Apply Deep Learning	- Cleaning Datasets, SVM, Deep Learning, error analysis across algorithms (F1 score, accuracy, precision, recall)

Describe Dataset (Q5**):**

We will use mask v. no mask image dataset provided by Kaggle, which are already separated into two folders(mask and no mask). We plan on adding features to the dataset by extracting facial landmarks (nose, mouth, jaw) from the images with dlib, OpenCV, and Python. The facial landmarks will be binary columns, where the value of 1 equates to that facial landmark being present in the image. We will combine this dataset with another one provided by Cornell that provides correct v. incorrect wear of a mask.

Timeline & Expected Deliverables (including Final Demo) (Q6+7**):**

Timeframe (Q6**)**	**Goals (**Q6**)**	**Expected Class Deliverables (start Nov.) (**Q7**)**
	- Determine ***feature se***t for detecting mask, no mask, incorrect mask. - Gather any other important studies around our topic	
Oct. 11th-17th-	- Data Preprocessing is done	

Oct. 18th-24th (Take Home Midterm)	- Combine all of the datasets - Determine multi-class versus binary classifications - EDA - Create a facial landmark features set for the dataframe (What are our base features for comparing mask v. no mask?)	
Oct. 25th-31st	- KNN_Classifier, compare against a multitude of neighbors + distance formulas - Have some error analysis done for model - Determine which features are most correlated to target	
Nov. 1st- 7th	**Mini-presentations in class start**	- Present current set of performance metrics against our baseline study, as well as the work done to achieve our 'feature set' for our mask/no mask/incorrect wear of mask dataset. Include the direction of our research in designing our own face-mask detection model.
Nov. 8th-14th	- Implement a Naive Bayes classifier - Conduct appropriate error analysis on it Naive Bayes - Implement CNN - Conduct error analysis on CNN - Implement SVM - Conduct error analysis on SVM - ****Try to combine the best models to create our own face-mask detection algorithm**	
Nov. 15th-21st	- Decision tree algorithm - Conduct appropriate error analysis on it Decision tree - Implement Deep Learning	

	- Conduct error analysis on Deep Learning	
Nov. 22th-Dec. 5th	- Wrapping up our findings - Start preparing for final deliverables (powerpoint, poster board, algorithm, visual to show our algorithm in action)	
Dec. 6th -Dec 12th **[Final Presentation]**	Present our final model as well as the findings from the models we tested above (KNN, CNN, Naive Bayes, Decision Trees, SVM, Deep Learning) and how each model's performance metrics compare against our baseline in the following studies: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7386450/pdf/main.pdf .	

Plan to Evaluate the Project (Q8**)**

We will use the performance metrics gathered on the SVM, KNN, CNN, Naive Bayes, Deep Learning, and Decision Trees Models to create our own face-mask detection model. After analyzing our face-mask detection model's performance metrics, we will compare it to our baseline research model's accuracy of 99.64% and use that to evaluate the effectiveness of our findings in the project (Loey, M et al., 2021).

References

Loey, M., Manogaran, G., Taha, M., & Khalifa, N. (2021). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement : journal of the International Measurement Confederation*, *167*, 108288. https://doi.org/10.1016/j.measurement.2020.108288