Imagine being able to detect blindness before it happened.

Millions of people suffer from [diabetic retinopathy](https://nei.nih.gov/health/diabetic/retinopathy), the leading cause of blindness among working aged adults. **Aravind Eye Hospital** in India hopes to detect and prevent this disease among people living in rural areas where medical screening is difficult to conduct. Successful entries in this competition will improve the hospital’s ability to identify potential patients. Further, the solutions will be spread to other Ophthalmologists through the [4th Asia Pacific Tele-Ophthalmology Society (APTOS) Symposium](https://www.kaggle.com/c/aptos2019-blindness-detection/overview/aptos-2019)

Currently, Aravind technicians travel to these rural areas to capture images and then rely on highly trained doctors to review the images and provide diagnosis. Their goal is to scale their efforts through technology; to gain the ability to automatically screen images for disease and provide information on how severe the condition may be.

In this synchronous Kernels-only competition, you'll build a machine learning model to speed up disease detection. You’ll work with thousands of images collected in rural areas to help identify diabetic retinopathy automatically. If successful, you will not only help to prevent lifelong blindness, but these models may be used to detect other sorts of diseases in the future, like glaucoma and macular degeneration.

Images have five possible ratings, 0,1,2,3,4.  Each image is characterized by a tuple *(e*,*e)*, which corresponds to its scores by *Rater A* (human) and *Rater B* (predicted).  The quadratic weighted kappa is calculated as follows. First, an N x N histogram matrix *O* is constructed, such that *O* corresponds to the number of images that received a rating *i* by*A* and a rating *j* by*B*. An *N-by-N*matrix of weights, *w*, is calculated based on the difference between raters' scores:

An *N-by-N* histogram matrix of expected ratings, *E*, is calculated, assuming that there is no correlation between rating scores.  This is calculated as the outer product between each rater's histogram vector of ratings, normalized such that *E* and *O* have the same sum.

To solve the problem, it was suggested to all the founders of the World Championship, due to the rarity of the disease, that the four classes be combined together to create a model that differentiates whether this person is sick or not.

Indeed, more than one model has been obtained with high accuracy, high quality, and stability in response

The same idea was invented to collect the 2023 data after more than one failed attempt to create a model that recognizes only the types of diseases, but the attempt led to failure. The same idea was created to collect all the disease data together and build a model.

One of the easiest, most accurate and fastest models was CNN, and its accuracy ranged between 90 to 93 percent.