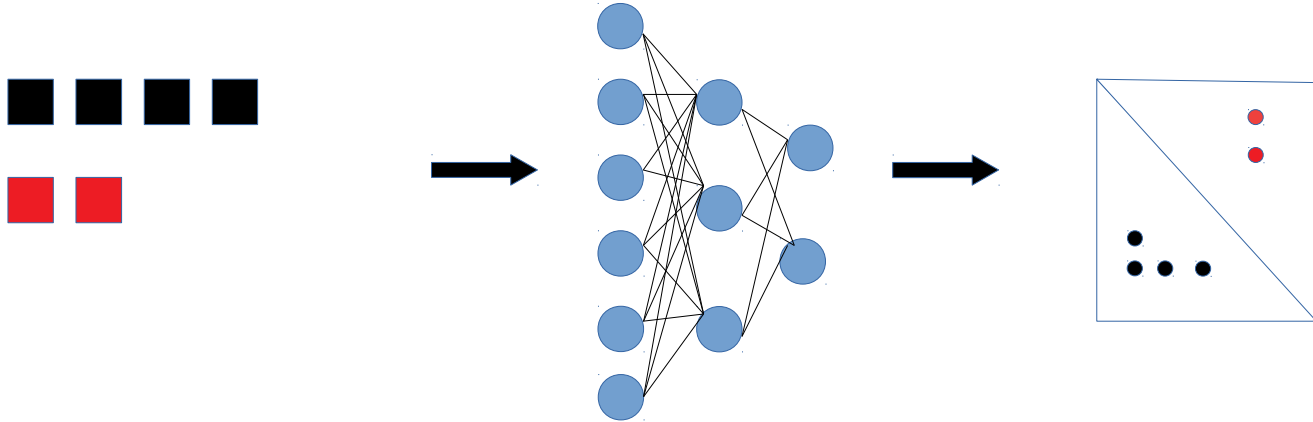


# Image Classification on Imbalanced Data Using Deep Learning



Work is available at

[https://github.com/shouvikcirca/PCB\\_Defect\\_Detection](https://github.com/shouvikcirca/PCB_Defect_Detection)

# Dataset

Initially, Balanced with 300X300 RGB images

Now, Imbalanced with 600X600 RGB images (Bigger holistically and instantially)

D (Defective)

ND (Non Defective)

Model	8 Fold Validation (Mean)		Test Set	
	D acc (%)	ND acc (%)	D acc (%)	ND acc (%)
Saswat Net	76.9	64.39	-	-
DN201 + ann	82.36	80	67	82
DN201 + conv	83.8	82.2	76	80
DN201 + conv2	84.727	81.3	83.05	74.64
Ensemble ( VGG16 + DN121 + DN201 )	75.84	72.5	66.1	87
DN201 + SVC	84.4	78.8	79.6	76.8

- DN -> DenseNet, VGG -> Visual Geometry Group
- DN and VGG are publicly available architectures
- The suffix number indicates the number of layers

Model	Description
DN201 + conv_i	Extracted features of Densenet + more conv layers + feedforward
DN201 + ann	Extracted features of Densenet + feedforward
Ensemble	Extracted features + ann( for 3 different models) + voting
DN201 + SVC	Extracted features + Support Vector Classifier

## Feature Extraction

- Pass the input through a pretrained network and use the output of the pretrained network as input for the downstream network that will be trained on your dataset.
- Pretrained networks are believed to have learned very good abstract features owing to large datasets they have been trained on.

## Ensemble

- Train multiple models
- Decide the label of a datapoint by majority of vote.

- Defective Accuracy =

Number of defective samples identified correctly

---

Total Number of defective samples

=

True Negatives

---

TrueNegatives + False Positives

- Non Defective Accuracy =

Number of non defective samples identified correctly

---

Total Number of non defective samples

=

True Positives

---

True Positives + False Negatives

### Saswat Net Fold Accuracies

# Fold	D acc (%)	ND acc (%)
1	100	0
2	77	83
3	100	0
4	0	100
5	81.35	89.83
6	77.9	93.2
7	100	0
8	79.6	84.7



# Proposed Work

Augment data by modelling distribution of data using Deep Generative models.



