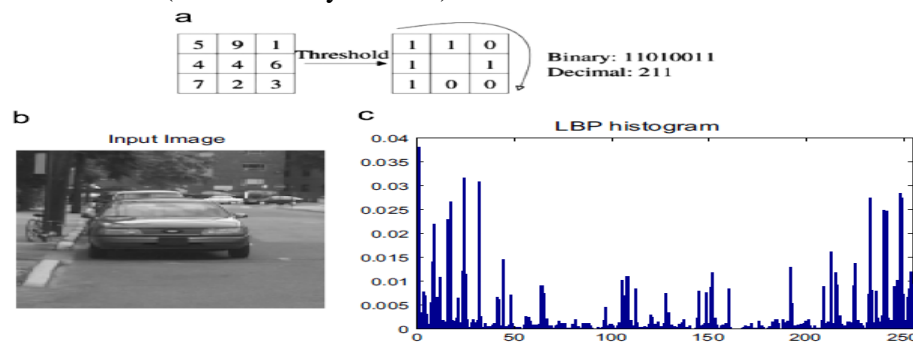


## TP5 Projets Apache Spark

Projet 1 : Développement d'un modèle de classification de gènes selon leurs implications à la Génodermatose (maladie génétique de la peau).

Utiliser Apache spark et les différents formalismes de classifications qu'on a vu (Arbre de décision, forêts aléatoires, réseau de neurones, classifieur Bayésien Naïf) afin de développer le meilleur modèle de classification pour prédire l'implication de gènes à la Génodermatose. Vous avez en pièce jointe une base de données réelles présentant 1290 gènes (dont l'implication/la non implication est confirmée) décrits par 224 variables descriptives (cliniques et biologiques).

Projet 2 : Développement d'un modèle de classification d'images en utilisant les histogrammes LBP (Local Binary Pattern).



Utiliser Apache spark et les différents formalismes de classifications qu'on a vu (Arbre de décision, forêts aléatoires, réseau de neurones, classifieur Bayésien Naïf) afin de développer le meilleur modèle de classification pour des images de voitures, de motos et de piétons. Les critères de paramétrages LBP seront :  $r=1$  (distance entre le centre et son voisinage), voisinage = 8, seuls les LBP uniformes seront retenus.

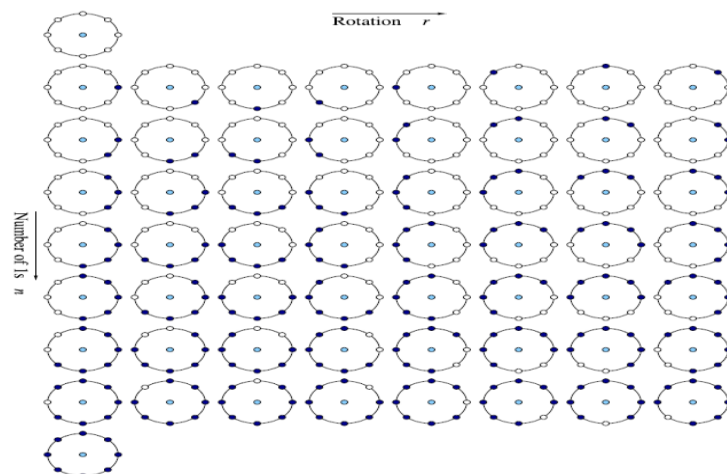


Fig. 2.4 The 58 different uniform patterns in  $(8, R)$  neighborhood