

Data Mining sur les images satellitaires pour la découverte des zones à risques d'épidémies

Boukaye Boubacar TRAORE
Doctorant – LGP Tarbes

Plan

- Introduction
- Méthodologie
- Etude de cas et Résultats
- Conclusion et perspective

I. Introduction(1)

II. Méthodologie

III. Etude de cas et Résultats

V. Conclusion et perspectives



Aires des vecteurs des épidémies

Risque d'épidémies



Climat et environnement



Aspect spatio-temporel des épidémies

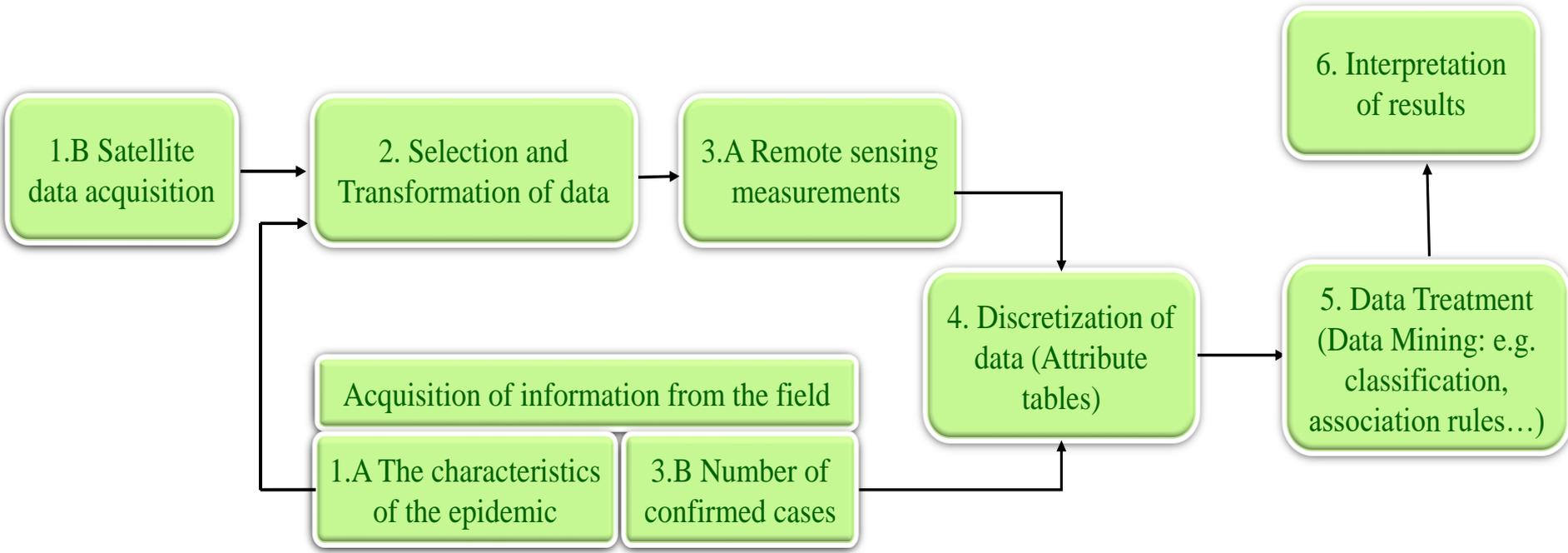


Télé-épidémiologie:

Les technologies spatiales au service de la lutte contre les épidémies transmises par l'eau, l'air ou des vecteurs en mettant en relation “*climat- environnement - santé*”.

[Bicout et al., 2015]

Méthodologie adoptée [Boukaye et al., 2017]



I. Introduction

II. Méthodologie

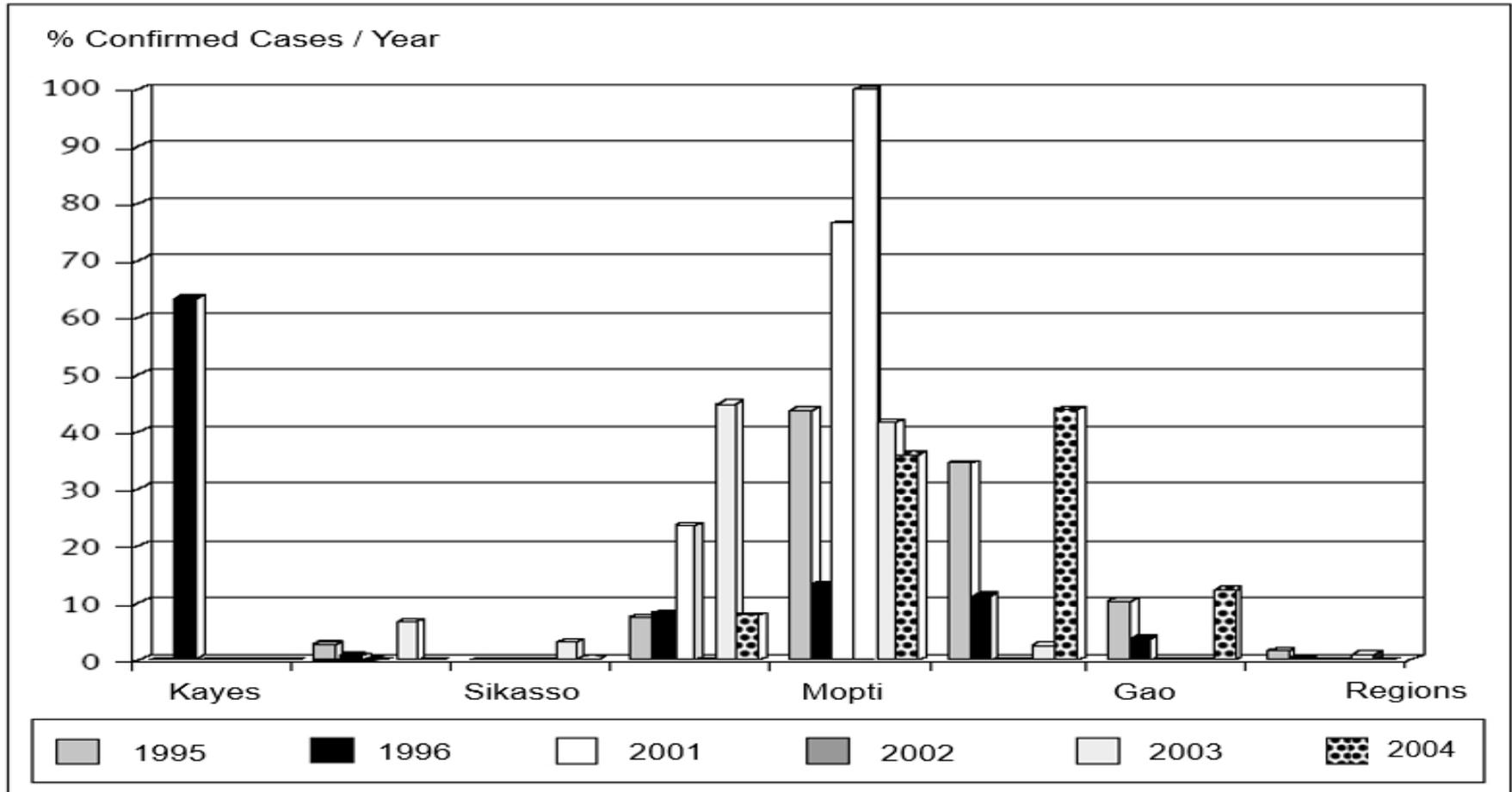
III. Etude de cas et Résultats (1)

V. Conclusion et perspectives

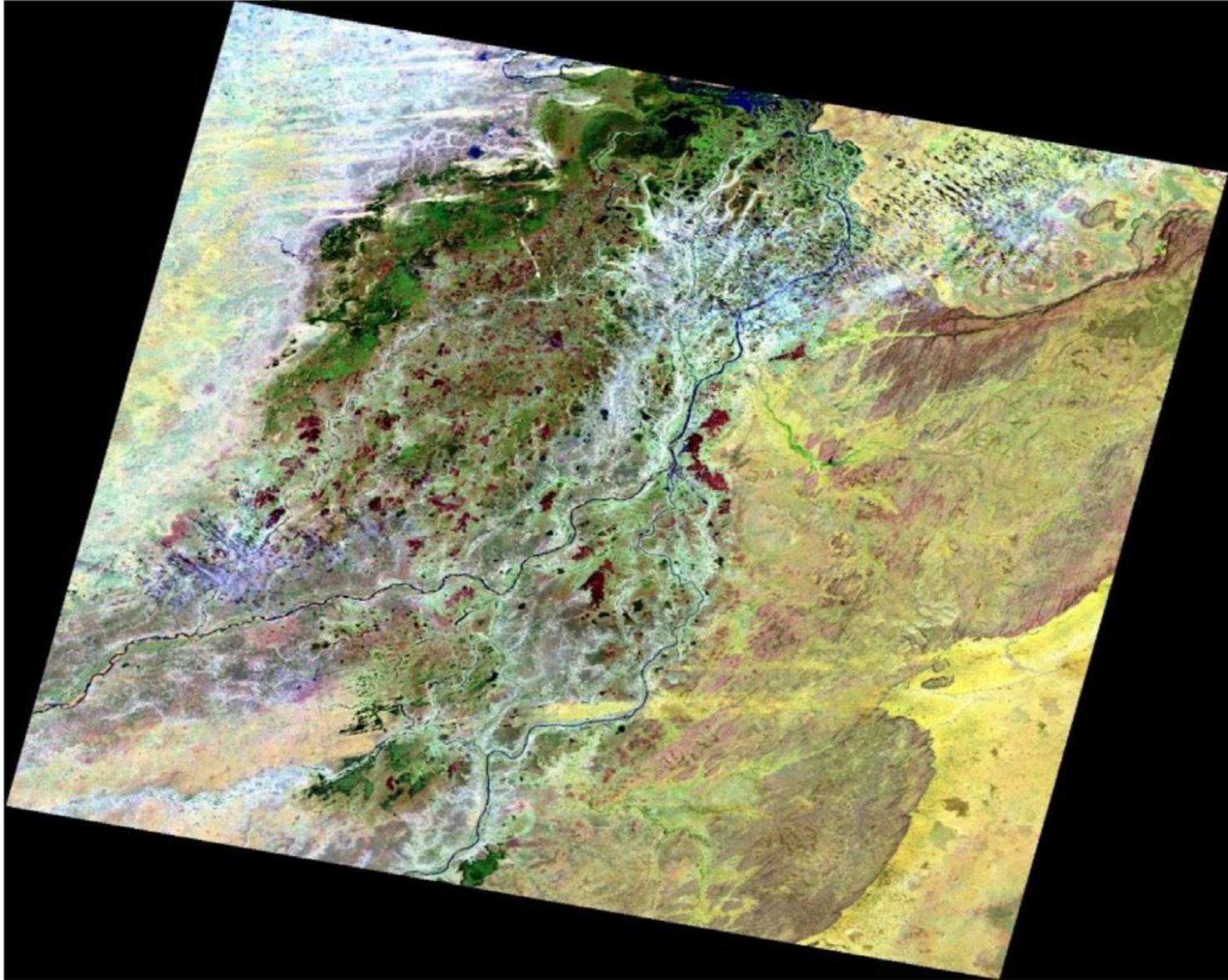


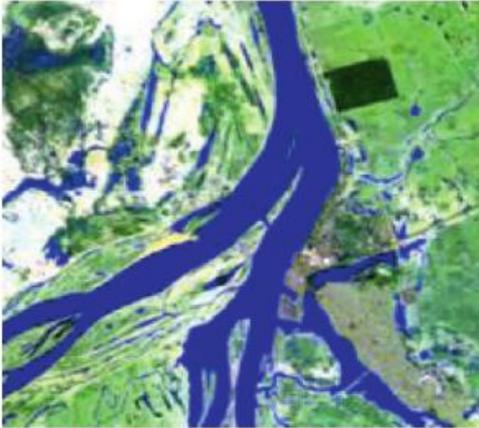
<http://www.msf.ch>

Répartition des cas de choléra au Mali de 1995 à 2004 selon les régions (Mopti la plus affectée)



Acquisition des images Landsat (Path = 197 and Row = 50)



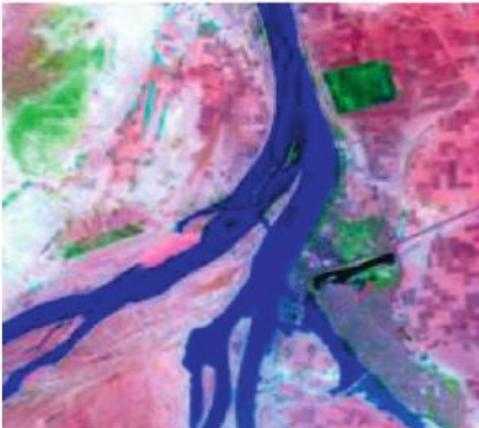


date: 1995/08/05
Landsat 4 Present

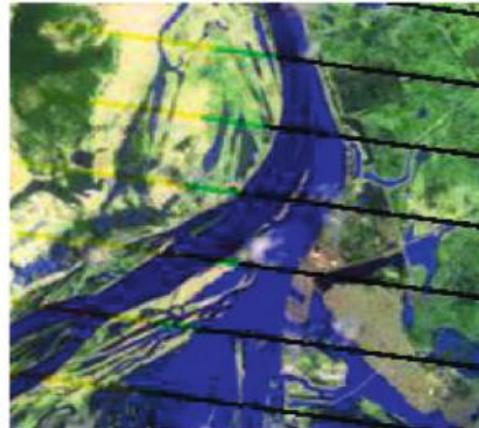
NO IMAGES
AVAILABLE IN
LANDSAT ARCHIVES
AT THIS AREA
FROM
1995/08/05-1998/02/18



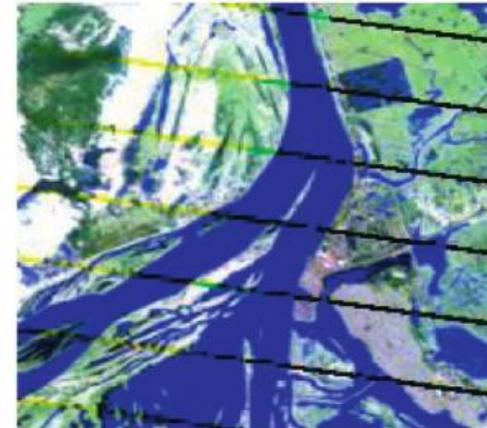
date: 2001/07/28
Landsat 7 SLC-on (1999-2003)



date: 2002/07/31
Landsat 7 SLC-on (1999-2003)



date: 2003/08/19
Landsat 7 SLC-off (2003->)



date: 2004/08/21
Landsat 7 SLC-off (2003->)



NO IMAGE AVAILABLE
IN LANDSAT ARCHIVES
AT THIS AREA
FROM
1995/08/05-1998/02/18



Date: 1995/08/05
Ground Truth (Percent)
Risk areas (Red): 17.24%
Not risk areas (Green): 82.76 %

Date: 2001/07/28
Ground Truth (Percent)
Risk areas (Red): 16.75%
Not risk areas (Green): 83.25 %



date: 2002/07/31
Ground Truth (Percent)
Risk areas (Red): 14.73%
Not risk areas (Green): 85.27%

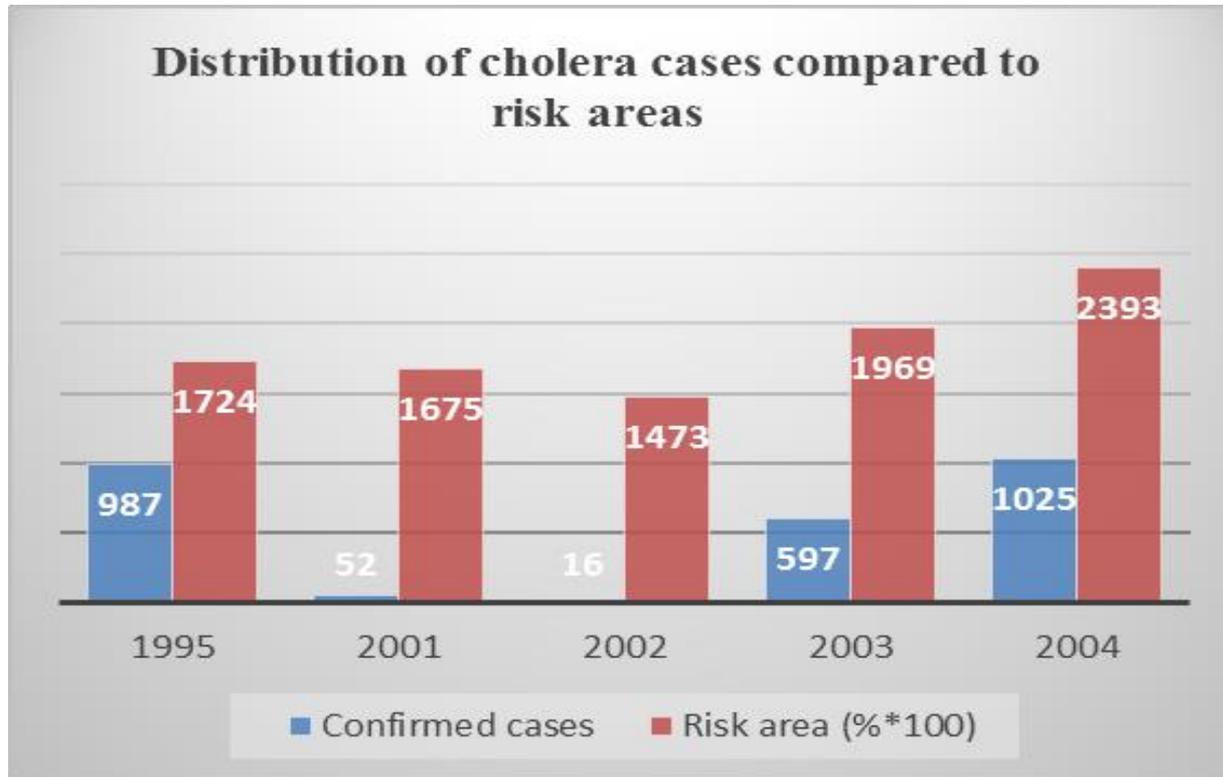


date: 2003/08/19
Ground Truth (Percent)
Risk areas (Red): 19.69%
Not risk areas (Green): 80.31 %



date: 2004/08/21
Ground Truth (Percent)
Risk areas (Red): 23.93%
Not risk areas (Green): 76.07%

Résultats de la classification et Données du terrain



Prétraitement des données pour l'algorithme de Data mining

Table 1: Système de codification de la BDD

Confirmed cases			Risk areas		
Code	Interval	Description	Code	Interval	Description
1	[0-350]	Low	11	[09%-16%]	Low
2	[351-700]	Medium	22	[17%-24%]	High
3	[701-1050]	High			

Table 2: BDD codifiée

Confirmed cases	Risk areas (%*100)
3	22
1	11
1	11
2	22
3	22

Processus de Data mining: règles d'association

Table 3: Résultats obtenus à partir de l'algorithme Apriori Association Rules [Agrawal & Srikant, 1994]

Association rules generated	Interpretation
1 \Rightarrow 11 #SUP: 2 #CONF: 1	If confirmed cases is low ([0-350]) Then risk area is also low ([09%-16%]) with a full confidence (100%)
22 \Rightarrow 3 #SUP: 2 #CONF: 0,66667	If the risk area is high ([17%-24%]) Then confirmed cases is high ([701-1050]) with a good confidence (66%)
3 \Rightarrow 22 #SUP: 2 #CONF: 1	If confirmed cases is high ([701-1050]) Then risk area is also high ([17%-24%]) with a full confidence (100%)

Post-traitement (Interprétation): à 66% le taux de contamination est lié au fleuve Niger

Conclusion

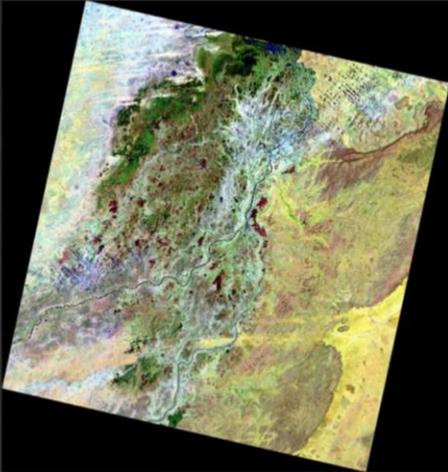


Perspective

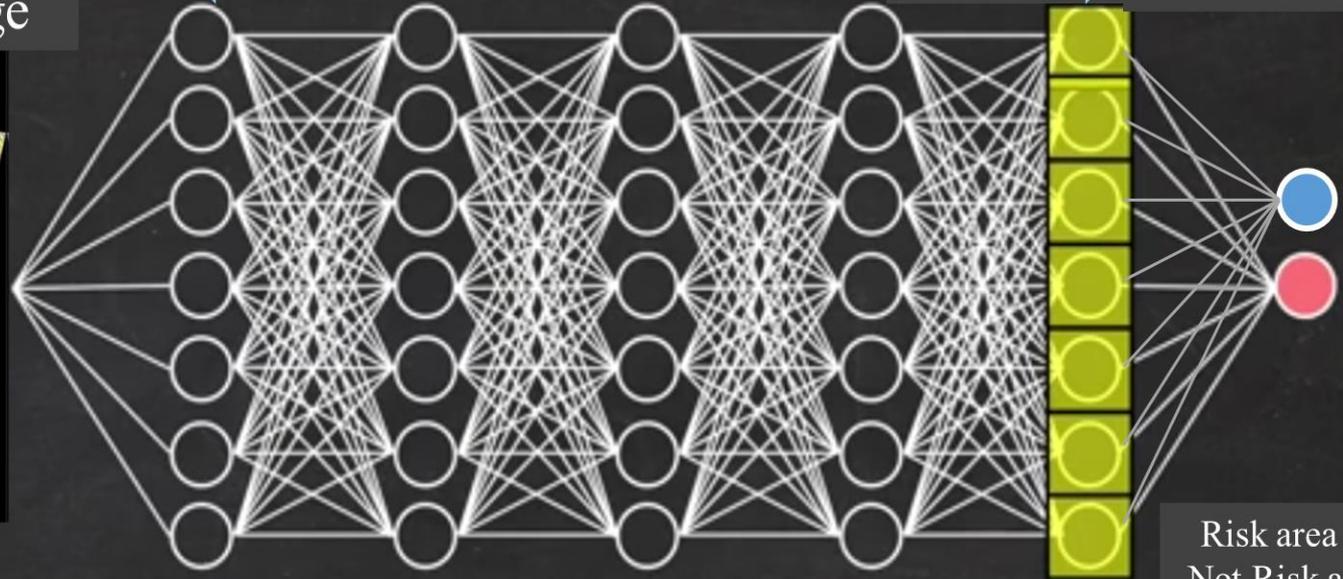


Perspective: *Deep Convolution Neural Network*

Input satellite image



Deep Learning



Classification

Risk area or Not Risk area

Merci pour votre attention

