

# **Approche connexionniste pour l'extraction de profils cas-témoins du cancer du Nasopharynx à partir des données issues d'une étude épidémiologique**

Khalid Benabdeslem\*, Mustapha Lebbah\*\*, Alexandre Aussem\* et Marilys Corbex\*\*\*

\*PRISMa, Université Lyon1,  
8, Avenue Niels Bohr, 69622 Villeurbanne Cedex, France  
{kbenabde, aaussem}@bat710.univ-lyon1.fr  
\*\* Université Paris 13, UFR de Santé,  
Médecine et Biologie Humaine (SMBH) - Léonard de Vinci- LIM&BIO  
74, rue Marcel Cachin 93017 Bobigny Cedex France  
lebbah@limbio-paris13.org,  
\*\*\*CIRC, Unité d'épidémiologie génétique  
150, cours Albert Thomas, 69280 Lyon Cedex 08, france  
corbex@iarc.fr.

**Résumé.** Dans cet article, nous présentons un système de découverte de connaissances à partir de données issues d'une étude épidémiologique cas-témoins du cancer du Nasopharynx (NPC). Ces données étant obtenues par une collecte de questionnaires, elles ont d'une part, la particularité d'être qualitatives et, d'autre part, de présenter des valeurs manquantes. Prenant en compte ces deux dernières contraintes, le système que nous proposons suit une démarche d'exploration de données qui consiste à (1) définir une procédure de codage des données qualitatives en présence de valeurs manquantes ; (2) étudier les propriétés de l'algorithme des cartes auto-organisatrices de Kohonen et son adaptation à ce type de données dans un cadre de découverte et de visualisation de groupes homogènes des cas cancer / non-cancer ; (3) post-traiter le résultat de cet algorithme par une classification automatique pour optimiser le nombre de groupes ainsi trouvés, et (4) donner une interprétation sémantique des profils extraits de chaque groupe. L'objectif général de cette étude est d'éclater le profil statistique global de la population étudiée en un ensemble de profils types (cancer ou non-cancer) et d'extraire pour chaque profil l'ensemble de variables explicatives du NPC à partir d'une cartographie bidimensionnelle.

## **1 Introduction**

L'algorithme des cartes auto-organisatrices de Kohonen, Kohonen (1994) représente un véritable outil de visualisation des données multidimensionnelles. Il permet de convertir des relations statistiques complexes et non-linéaires entre les données de grande dimension en une simple relation géométrique sur une topologie réduite. Cet algorithme permet de compresser