1.1 تمهيد

يعد علم التنقيب فى البيانات(Data Mining) من العلوم المهمة والنشطة الحديثة فى مجال علوم الحاسوب والتى شغلت اهتمام كبير فى مؤسسات البرمجة والإحصاء ومدراء الشركات فى العالم كما يعد جمع البيانات من الأشياء الضرورية لعملية التنقيب والتى من خلالها تستخرج النتائج المجهولة او التى لم تكن معروفة من قبل، فسنتعرف الى المقصود بتنقيب البيانات والفائدة من تطبيقة فى المجالات المختلفة واوجة القصور في التنقيب وفوائده وغيرها مما يتعلق بماهية هذا العلم سنتعرض له اشاءالله فى هذا الباب.

تجيب تقنيات التنقيب علي العديد من الأسئلة، وفي وقت قياسي ، خاصة تلك النوعية من الأسئلة التي كان من الصعب الإجابة عليها، إن لم يكن مستحيلاً ، وإن وجدت فإنها تستغرق وقتاً طويلاً.

وقد نشأ علم التنقيب في قواعد البيانات كنتيجة طبيعية للنمو الكبير فى أحجام قواعد البيانات وصفحات الويب خاصة بعد الإنتشار الواسع لإستخدام أنظمة المعلومات وتراكم الكم الهائل من البيانات التي أصبحت متداولة يومياً في العديد من المجالات، الأمر الذي جعل المتخصصين يفكرون فى بناء نظم تثير تحور هذه البيانات وإستخلاص بعض المعلومات الخفية منها والتي لايمكن الوصول اليها بالاسئلة التقليدية التى تستخدم مع قواعد البيانات.

2.1 مشكلة البحث

توجد بيانات مرضى ضخمة فى نظام تسجيل مرضى السرطان بمستشفى الذرة بالخرطوم وفى كثير من الحالات توجد مجموعات غير معروفة للعاملين والخبراء والاطباء بالمستشفى ومن هنا سوف يحاول البحث ان يستخدم طريقة تحليل العنقدة اوالتجمعات للبيانات(clusters) لهذه البيانات للوصول الى مجموعات جديدة لم تكن معروفة من قبل ولم يهتم بها الخبراء والعاملين فى المجال الطبى.

3.1 أهداف البحث

تطبيق تقنيات إستخلاص المعرفة الطبية علي بيانات مرضى السرطان للتوصل إلى بعض القواعد و تقسيم البيانات الرقمية الطبية إلى عدة عناقيد اوتجمعات وذلك بغرض الوصول الى عناقيد غير معروفة للعاملين بالحقل ومفيدة لهم.

4.1 منهجية البحث

يتبع البحث المنهج التطبيقى التحليلي.

5.1 هيكل البحث

تم تقسيم هذا البحث الى عدد من الابواب تناولت مفردات هذا البحث الاتى :

- الباب الأول تناول تعريف مشكلة البحث و أهداف البحث منهجية البحث وهيكل البحث.
- الباب الثانى تناول خطوات تنقيب البيانات (Data Mining) فى قواعد البيانات واوجة القصور فى تنقيب البيانات ثم الأدوات والبرامج المستخدمة فى تنقيب البيانات ايضا تناول طرق تنقيب البيانات والمنهجية المستخدمة فى التنقيب والخوارزميات المستخدمة فى عملية التنقيب بنوع من الاختصار وتناول البحث الخوارزمية المستخدمة بنوع من التفصيل.
- الباب الثالث تحدث عن الجزء العملى (Case Study) الذى صاحب البحث والتقارير التى تم توصل اليها عمليا .وتناول ايضا النتائج
- الباب الرابع تناول التوصيات التى خرج بها هذا البحث والخاتمة.

التنقـيـب في البيـــانات

التنقيب في البيانات

Data Mining

1.2 تعريف

التنقيب في البيانات هي عملية بحث محوسب عن معلومات ومعرفة من مجموعة بيانات ضخمة دون فرضيات مسبقة. هناك هدفان أساسيان للتنقيب في البيانات وهما التنقيب الاستشرافي والتنقيب الوصفي. التنقيب الاستشرافي ينتج عنه نموذج عن النظام الذي تصفه البيانات المستخدمة في التنقيب. أما التنقيب الوصفي فينتج معلومات جديدة بناء على المعلومات الموجودة داخل البيانات المستخدمة في عملية التنقيب [1]. وتوجد عمليات أساسية تعتبر ضمن عمليات أو مهام التنقيب هي:

أ- التصنيف

ب- الانحدار

ت- العنقدة (التجميع)

ث- التلخيص

ج- نماذج الاعتماد

ح- الكشف عن التغيرات والانحرافات .

وهناك ثلاثة أهداف للتنقيب في البيانات:

أ- من أجل تعليل بعض الظواهر المرئية. مثال: لماذا زادت نسبة المدخنين في الوطن العربي؟.

ب- من أجل التثبت من نظرية ما. مثال: التثبت من النظرية التي تقول بأن الأسر الكبيرة تهتم بالضمان الصحي أكثر من الأسر الصغيرة عددا.

ت- من أجل تحليل البيانات للحصول على علاقات جديدة وغير متوقعة. مثال: كيف سيكون الانفاق العام إن كان ملازما لعمليات خداع واسعة من قبل البطاقات الائتمانية[1].

ث- وسائل التنقيب في البيانات: هناك عدة وسائل مختلفة من أجل التنقيب في البيانات. اختيار الوسيلة المناسبة يعتمد على طبيعة البيانات تحت الدراسة وعلى حجمها. يمكن اجراء عملية التنقيب في البيانات بالمقارنة مع سوق البيانات ومخزن البيانات[1]. وتوجد بعض من هذه الوسائل هي

- التفكير واستخلاص النتائج والقوانين من أمثلة حية (Case-Based reasoning)
- الكشف عن قانون (Rule Discovery): مثل البحث عن منوال معين أو علاقة معينة في جزئية كبيرة من البيانات
- معالجة الاشارات (Signal Processing): مثل ايجاد الظواهر المتشابه مع بعضها البعض.
- الشبكات نووية (Neural Nets): مثل تطوير نماذج قابلة لتنبؤ النتائج. هذه النماذج تم تطويرها بناءا على أسس تم استنباطها من عقل الانسان.
- منحنيات غير ثابتة (Fractals): تصغير البيانات الكبيرة من دون ضياع المعلومات.

وسائل التنقيب في البيانات تُستعمل وبنجاح في الكثير من التطبيقات الحقيقة حول العالم. التطبيقات التالية تشمل بعضا من الأمثلة:

أ- كتابة تقرير مختصر عن فئة معينة (Profiling Populations): مثل تطوير وإنشاء تقارير موجزة عن الزبائن المهمين وعن بطاقات الائتمان.

ب- تحليل النزعة التجارية (Analysis of Business Trend): مثل ايجاد الأسواق ذات قدرات النمو القوية أو الضعيفة.

ت- التسويق لفئة معينة (Target Marketing): مثل ايجاد الزبائن من أجل منح التخفيضات لهم لسبب معين.

ث- تحليل الاستعمال (Usage Analysis): ايجاد منوال معين لإستعمال الخدمات والسلع.

ج- فعالية الحملة (Campaign Effectiveness): مقارنة استراتيجيات الحملات مع بعضها البعض من أجل ايجاد أكثرها فعالية وتأثيرا.

ح- جاذبية السلعة: ايجاد السلع التي تباع مع بعضها البعض.

2.2 نماذج تطبيقية في تنقيب البيانات

نظرا للمزايا التي يمتلكها هذا الاتجاه تم تنفيذ بعض من اساليبه على سبيل التجارب التطبيقية في موضوعات علم المعلومات ومنها تجربة بناء مكنز آلي وهو عبارة عن نظام لجمع المصطلحات العلمية التى تساعد فى عمبية البحوث العلمية للباحثين وباستخدام اسلوبي قاعدة الارتباط (Rule Association) والعنقدة (clustering) إذ وبواسطة الاسلوب الاول تم تحليل مستخلصات بحوث علمية وتحديد مجموعة المصطلحات الكبيرة (large Item set) وبواسطة الاسلوب الثاني تم تجميع المصطلحات في عناقيد ومنها تم الوصول الى المصطلحات العريضة والمصطلحات الضيقة والمصطلحات المترابطة وبالنتيجة تم الحصول على مكنز آلي باستخدام اساليب استخراج البيانات اما التجربة الثانية فكانت تدور حول استخدام اسلوب وخوارزميات التصنيف المستخدمة في استخراج البيانات لغرض ايجاد طريقة جديدة في التصنيف الآلي لمصادر المعلومات في المكتبات، في هذه التجربة تم تحليل مجموعة من البحوث في موضوع تكنولوجيا المعلومات وباستخدام خوارزمية التصنيف (Classification) تم الحصول على اصناف محددة سيتم اتباعها في المستقبل [5].

ولازال موضوع تطبيق استخراج البيانات في مجال علم المعلومات وتحديدا استرجاع المعلومات ارضا بكرا بحاجة الى المزيد من البحث واجراء التجارب لغرض الحصول على العديد من الموضوعات والافكار الجديدة التي من شأنها الارتقاء بمستوى خدمات المعلومات .

بدأت تشهد المعلوماتية ازدياداً كبيراً في الأنشطة الاقتصادية في الشركات الكبرى وبدأت مفاهيم جديدة تظهر معتمدة على هذه التقنيات فمثلاً (Web Mining) الذي يجمع بين مخازن المعلومات وشبكة الانترنت والغاية منه إنشاء بنك معلومات عالمي متعدد اللغات مختلف المواقع ويمتاز بسهولة استعماله والـ (Text Mining) ونعني بها استخراج المعرفة المفيدة والغير ظاهرة في كميات كبيرة من النصوص الغير منتظمة وبتعبير آخر تحويل النص الحر إلى نص يمكن تحليله آلياً ويعتبر الـ (Text Mining) من المجالات الجديدة في البحث وخاصة مع انتشار الانترنت بحيث يحاول إيجاد حلول للوصول إلى المعرفة في الكم الهائل من النصوص الموجودة على شبكة الانترنت بإستخدام تقنيات الفهم الآلي للغات الطبيعية فتتم هذه العملية من خلال معالجة أولية للنص وذلك بإستخراج الكلمات والمفاهيم وبعد ذلك من خلال إيجاد العلاقات بين المفاهيم وتمثيل النص في قواعد ربط متعدد

الوسائط (Data Media) والذي يضم إضافة إلى النص الصورة والصوت والخرائط بأنواعها وإمكانية الاسترجاع في محتوى أي من الوسائط المخزنة فيه وتستعمل فيه تقنيات معقدة في التوثيق والبحث وتستخدم أدوات الذكاء الاصطناعي والتعرف على الأشكال.[6]

3.2 التحديات التى تواجه عملية التنقيب

بالرغم من ان المنتجات التى تقوم بتنقيب البيانات تعتبر قوية جدا ، الا انها لاتفى بالغرض لوحدها، وحتى تتمكن هذه المنتجات من النجاح فى التنقيب، تحتاج الى مهارات تقنية، ومحللين متخصصين لديهم القدرة على التحليل الجيد، وترجمة النتائج بصورة تفيد المؤسسة وتجعل عملية التنقيب فعالة، وبالتالى نجد ان التحديات التي تواجه عملية التنقيب تتمثل فى الاتي :

1) عملية التنقيب تساعد فى اكتشاف الانماط والعلاقات الا انها لاتخبر المستخدم بقيمة او اهمية النمط او العلاقة التى استنبطتها ، لان هذة هى مهمة المستخدم ، فالالة المستخدمة فى استخراج الذهب لاتريك كيف تقوم بتشكيلة. وصلاحية النمط او العلاقة المكتشفة تعتمد على امكانية مقارنتها مع ظروف العالم الحقيقى فقد لا نستطيع تطبيق ما اخرجتة أداة التنقيب على ارض الواقع.

2) عملية التنقيب تعرف الارتباطات بين المتغيرات والسلوك الا انها لاتعرف العلاقات الغير رسمية.

4.2 الأدوات والبرامج المستخدمة فى تنقيب البيانات

هناك العديد من المنتجات والادوات والبرامج التى تقوم بعملية تنقيب البيانات والتى تنتج عن شركات كبيرة ومعروفة مثل(Oracle، SPSS ،IBM،Microsoft)

وغيرها من الشركات المعروفة ، ومن هذه البرامج التى تقوم بعملية تنقيب البيانات واستخدام الخوارزميات المطبقة التى تقوم بعملية التنقيب هو مايعرف بـ:

Clementine 7.0 1.4.2

كما ذكر فى دليل المستخدم لهذه الاداه فهى أداة ناضجة للتنقيب في البيانات و التي تهدف إلى إتاحة خبراء المجال (المستخدمين العاديين) للقيام بتعدين البيانات الخاصة بها.

قامت بانتاجة شركة (SPSS) به مكتبات كاملة لعمل تنقيب البيانات والتى تعمل على استخدام التنقيب بواسطة التصنيف (Classification) والتحليل العنقودى(Clustering Analysis) وقواعد اكتشاف الإرتباطات والعلاقات (Associations Rule) ويمكن من خلالة استخدام اى نوع من هياكل البيانات مثل الملفات بانواعها، قواعد البيانات، الجداول وغيرها. كما أنه يستخدم أشكال الرسومات البيانية لتوضيح نتائج التنقيب ويعتبر من أبرز وأحدث البرامج فى تنقيب البيانات .

Weka 2.4.2

(Weka) وهى عبارة عن أداة تحتوى على مجموعة من خوارزميات تعلم الاله تم انشاءها وتطويرها من قبل جامعة(Waikato) بنيوزلاند ويعد من المصادر المفتوحة (open source software) حيث يقوم بعمليات تنقيب البيانات ويحتوى على أداة لتجهيز البيانات والتصنيف والعنقدة وقواعد الارتباط وعمليات التمثيل الاحصائى لهذه الخوارزميات .

5.2 خطوات إكتشاف المعرفة فى قواعد البيانات

هنالك خطوات قياسية ومتعارف عليها ذكرت فى بعض الاوراق العلمية مثل Knowledge Discovery in Database(KDD) وذلك لإكتشاف المعرفة وهى:

أ- تحديد المشكلة (Problem specification)

والغرض من هذه المرحلة هو تحويل المشكلة إلى تعريف محكم ومحدد للمشكلة والعمليات التي يتم تنفيذها في هذه المرحلة هى الدراسة الأولية للمشكلة ومحاولة ايجاد تجمعات اوعناقيد شاذه غير معروفة

ب. الموارد (Resource)

وهي قائمة من متطلبات الموارد،ويتم تحديدها في مرحلة تحديد المشكلة وتضم الخوارزميات المستخدمة في التنقيب وكل الاشياء التى نحتاجها لانجاز عملية التنقيب من اجهزه وبرمجيات. وتحتوى على الخطوات الاتية:

أ- تجميع البيانات

هنا يتم تحديد البيانات التى يتم التنقيب فيها وتحديد خصائص تلك المصادر يتم فى تقرير يشتمل على مجموعة من التعليمات مثل مصدر البيانات، المالك، مدير قاعدة البيانات.

ب- وصف البيانات

ويشتمل على وصف لكل ملف اوجدول فى قاعدة البيانات وهناك بعض الخصائص التى توثق فى تقرير وصف البيانات منها:

1- عدد الحقول

2- عدد السجلات التى تحوى قيم مفقودة

3- أسماء الحقول

ج- اختيار جزء قاعدة البيانات الذى سيتم التنقيب فيه

وهذه الخطوة لاتعنى ان نأخذ عينة من قاعدة البيانات او نختار مجموعة من البيانات بشروط معينة وانما هى عملية ازالة تامة للبيانات الغير مطلوبة او التى ليست لها علاقة بالمشاكل المحددة.

ج. تنقية البيانات(Data cleaning)

وهي مرحلة عزل البيانات التي تحتوي على التشويش أو الشوائب من مجموعة من البيانات وأيضا يتم في هذه المرحلة معالجة البيانات المفقود وتقيم نوعية البيانات فهناك مشاكل كثيرة تظهر فى نوعية البيانات ، فقد يكون هناك حقل ادخلت له قيمة خاطئة مثل إدخال حقل اسم الطبيب فى حقل اسم المريض اوقد تكون هناك حقول هى بمفردها تحتوى على قيم صحيحة لكن عندما ندمجها مع حقول اخرى تبدو منطقيا خاطئة فيجب ان ننتبة الى البيانات خاصة عند جلبها من مصادرمختلفة.

وايضا البيانات المفقودة قد تسبب مشاكل لاتبدو جلية للوهلة الاولى فمثلا اذا تم إلغاء كل سجل به حقل فارغ فنحصل فى النهاية على قاعدة بيانات صغيرة جدا او صورة غير دقيقة لقاعدة البيانات. وفى الواقع هذه القيم المفقودة قد تشكل أهمية فى حد ذاتها فمثلا يمكن لبعض الاشخاص ان يتركوا حقل الدخل فارغا فى حين انك قد تحتاجة فى بناء نموذج تنقيب البيانات.

ويجب الاخذ فى الاعتبار انه ليس بالامكان معالجة جميع المشاكل فمن الافضل ان تقوم بعمل اجراءات وعمليات مراقبة لتفادى مشاكل نوعية البيانات.

د. المعالجة التمهيدية (Pre-processing)

تعتبر هذه المرحلة هي أول مرحلة فعلية في المشروع ويمكن تنفيذها أكثر من مرة أثناء سير المشروع ، وقد تحسن في جودة النتائج المتحصل عليها في مرحلة تنقيب البيانات. ومن طرق المعالجة التمهيدية مايعرف بتكامل البيانات فالبيانات التى يحتاج اليها قد تكون موجودة فى قاعدة بيانات واحدة او فى عدة قواعد بيانات.

تكامل البيانات يعنى تجميع البيانات فى قاعدة بيانات واحدة للتنقيب فيها ، وعند تكامل البيانات من المصادر المختلفة يجب ان ننتبه الى الاختلافات التى قد تكون فى قيم البيانات لانها سبب أساسى فى مشاكل نوعية البيانات، فهناك اختلافات جوهرية بين الموسسات فى طريقة تعريف البيانات او فى استخدام انظمة ادارة قواعد بيانات مختلفة . فى بعض الاحيان قد يكون هناك تضارب فى البيانات خاصة لوكانت البيانات مجتمعه من بلدان مختلفة فمثلا الدولار الامريكى والدولار الكندى لايمكن اضافتهما من غير تحويل [6] .

كما يعد قاموس البيانات(Data Dictionary) مهم جدا فالمعلومات التى يتم الحصول عليها تعتبر أساس البنيه التحتية لقاموس البيانات،وبشكل جوهرى يعتبر قاموس البيانات هو قاعدة بيانات لقاعدة البيانات نفسها.حيث انها تمنح معلومات تساعد فى التصميم الفيزيائى لقاعدة البيانات بالاضافة لذلك تساعد المحللين فى فهم البيانات وبناء النموذج.

د. توحيد البيانات (Data Integration)

في هذه المرحلة غالبا ما تكون مصادر معالجة البيانات متغيرة العناصر وربما تكون مجتمعة في مصدر شائع، مثلا تكون هناك اكثر من قاعدة بيانات بأشكال مختلفة فنجد بيانات فى هيئة اوراكل او اكسس او اكسل وغيرها فلابد من توحيد هذه المصادر فى جهة واحدة .

ه. تنقيب البيانات(Data Mining)

في هذه المرحلة يتم تطبيق أسلوب ذكي لاستخراج معرفة مفيدة حيث يتم استخدام بعض الخطوات وتطبيق خوارزميات لاستخراج نتائج مفيدة قدر الامكان. هذه الخطوة يتم تكرارها حتى الوصول للنتائج المرجوه ويجب إستكشاف النماذج البديلة حتي يتم الحصول على أكثرهم فائدة في حل المشكلة. وبعد إختيار طريقة التنقيب التنبؤية مثلا(طريقة التصنيف classification أو طريقة الارتداد regression) يجب إختيار النموذج الذي يصنع التنبؤ وذلك إما أن يكون شجرة قرار (decision tree) ، الشبكة العصبية (neural network) أو غيرهم .

اما عملية بناء نموذج التنقيب فى العنقدة فتتطلب تحديد الحقول التى يجب ان تشارك فى المجموعات ومن ثم لا يتم تحديد حقل اخراج كما فى التصنيف ولابد من تحديد نوعية البيانات المشاركة فى العنقدة وايضا يتطلب تحديد احدى خوارزميات العنقدة مثل ال (K-means)او(Kohean) او(Two-step) اوغيرها من الخوارزميات . ومن ثم ينتج نموذج يمكن تحديد اى عدد من العناقيد فيه وبالملاحظة والرجوع الى خبير فى مجال البيانات المجمعة تكون النتائج جيدة.

و. تقييم النتائج (Evaluation of Results)

ويكون التقييم في الأتي:

1. التنفيذ في قاعدة بيانات أولية.
2. التبسيط.
3. التعميم.

وبعد استخراج النتائج المهمة والتي تمثل المعرفة يتم تقييمها استنادا إلى مقاييس محددة.

س. تفسير النتائج (Interpretation of Results)

تتضمن هذه المرحلة ترجمة المعرفة المكتشفة كذلك مشاهدة للمعرفة المستخلصة والمشاهدة تزيد من القدرة على الفهم فى منظور الإنسان ،ويستطيع الشخص تقييم المعرفة المنقبة بصورة اوتوماتيكية أو نصف اوتوماتيكية لتحديد الاهتمام الحقيقي والمفيد للمستخدم.وايضا تحرى واستكشاف البيانات مهم فالهدف الأساسى من التحرى هو تميز الحقول الأكثر اهميه فى توقع النتائج وتقرير اى القيم المشتقة قد يكون مفيدا. مشروعية النموذج لايعني بالضرورة صحته ، أحد الأسباب الرئيسية لهذه المشكلة وجود افتراضات ضمنية داخل النموذج ، لذا فإنه من الضروري اختبار النموذج في العالم الحقيقي ، فإذا كان النموذج يستخدم في الوصف لمخاطر مرض السرطان مثلا، لابد من تجريبه علي مجموعة البيانات التى

تنتج ، لأن الخطر الأكبر يتمثل في النماذج الغير صحيحة ومن الأهمية أن تقوم بتجربة لتتحقق من نتائج النموذج الذي بنيته.

فى مجموعات البيانات التى تحتوى على المئات او ألاف من الحقول فقد يبدو أكتشاف البيانات بمثابة مضيعة للزمن وعملية مرهقة جدا . فالاستجابة السريعة للحاسوب المستخدم يشكل عاملا مهما فى هذه المرحلة حيث ان طبيعة التحرى والأستكشاف الذى تجرية على البيانات قد يختلف تماما اذا كنت ستنتظر بعض الزمن لرؤية بعض الرسومات البيانية فكيف يكون اذا كان هذا الانتظار يوما كاملا.

ح. تمثيل المعرفة (Exploitation of Results)

وهي المرحلة الأخيرة من مراحل اكتشاف المعرفة في قواعد البيانات وهي المرحلة التي يراها المستفيد،ويستخدم في هذه المرحلة الأسلوب المرئي لمساعدة المستفيد في فهم وتفسير النتائج.

6.2 أهم طرق تنقيب البيانات

1.6.2 قاعدة الارتباط (Association Rule)

قواعد الارتباط (Associations Rule) هي احدى الواجهات الواعدة من الـ (Data Mining) كاداة من ادوات اكتشاف المعرفة (KDD) ولديها القدرة على تصفح كميات هائلة من البيانات ، وهي تسمح بالتقاط كل القوانين الممكنة التي تشرح بعض الصفات الموجودة اعتمادا على وجود الصفات الاخرى. وبمعنى اخر هي قواعد ارتباطية معينة بين مجموعة من البيانات في قاعدة البيانات [7].

2.6.2 التصنيف (Classification)

يستخدم التصنيف بشكل واسع في حل الكثير من المشكلات خاصة تلك التي تتعلق بالاعمال (Business)من خلال تحليل مجموعة من البيانات ووضعها على شكل اصناف او اقسام يمكن استخدامها فيما بعد لتصنيف البيانات المستقبلية [2].

3.6.2 العنقده (Clustering)

1.3.6.2 تعريف العنقدة

وهي عملية تقسيم البيانات الى مجموعة من الأصناف اعتمادا على اشتراكها بالخواص المتشابهة وان العنقدة هي تقسيم غير موجه للبيانات [2].

ومن اهم الطرق المستخدمة فى العنقدة هى العنقدة الهرمية وهى تعمل هذه الطريقة على تجميع البيانات بإفتراض ان كل عنصر من العناصر الموجودة فى البيانات عبارة عن عنقود لوحدة ومن ثم تقوم هذه الطريقة بقياس التشابة بين اقرب عنقودين إعتماداً على المسافة حيث تحث عملية دمج للعناصر المتشابهة مع بعضها البعض [3]. ومن أبرز امثلتها خوارزمية الخطوتين(Two-Steps).

اما الطريقة الثانية فهى العنقدة الجزئية وهى من أهم طرق العنقدة وهذه الطريقة تفترض ان كل العناصر الموجودة عبارة عن مجموعة واحدة ومن ثم تختار عنصرين بصورة عشوائية وتحاول قياس التشابه بين العنصرين وأكبر العناصر تشابها هى التى تنقسم وهكذا تتوالي الانقسامات الى أن تصل لعدد الانقسامات المطلوبة، ومن أشهر أمثلتها هى خوارزمية (k_means) [3] .

2.3.12.2 خوارزمية الـ (Two-Step)

هى واحدة من خوارزميات العنقدة التى تعمل على تنقيب البيانات ، وجاء إسماها(Two-Step) نظرا لانها تقوم بعملية العنقدة من خلال خطوتين او مرحلتين. المرحلة الاولى هى أن الخوارزمية تمر على كل العناصر الموجودة وتقوم بعملية تجمعات اوعنقدة صغيرة لهذه العناصر، اما المرحلة الثانية تقوم الخوارزمية بعملية دمج العناصر القريبة مع بعضها البعض بواسطة قياس البعد بين العناصر حتى تصل الى عدد المجموعات الذى يطلبه المستخدم.

ميزة الخوارزمية انها يمكن ان تعطى نتائج عنقدة من غير ان يحدد المستخدم عدد العناقيد او المجموعات التى ستنتج ولكنها تاخذ زمن مقارنة بالخوارزميات الاخرى.

3.3.12.2 كيفية عمل خوارزميةالـ Two-Stepsرياضيا فى قياس المسافة لاجراء عملية الدمج

عمل الخوارزمية رياضيا فى قياس المسافة لإجراء عملية الدمج يعتمد على المتغيرات فالمتغيرات اما ان تكون متقطعة اوتصنيفية (Categorical variable) وهي المتغيرات التي تكون قيمها صغيرة فإذا كان الحقل لديه احتمالان لقيمته إما نعم او لا او اما ذكر او انثى كما فى حقل نوع المريض فهنا نقول أن الحقل ثنائي(binary) أو يكون هنالك أكثر من احتمالان لقيمة الحقل فمثلاً تكون القيم لحقل نوع اورام السرطان هي : خبيث ، حميد، وبين بين اى بين الخبيث والحميد فهنا نقول أن الحقل متعدد (Multiclass) . او تكون متغيرات مستمرة اومتغيرات رقمية (Numerical or continuous Attribute) وهي القيم التي تؤخذ من عدد كبير من القيم والتي يكون لديها ترتيب (يمكن ترتيبها) مثل حقل رقم المريض ، فالمسافة بين العنقودين فى حالة المتغيرات المتقطعة تستخدم مايعرف بال(Log-likelihood distance) [8] وهى عبارة عن ثلاث معادلات ، يمكن التعرف عليها كالاتي:-

$$d(j,s)=\xi_j+\xi_s-\xi_{<j,s>} \quad (1)$$

$$\xi_v=-N_v\left(\sum_{k=1}^{k^A}\frac{1}{2}\log(\hat{\sigma}_k^2+\hat{\sigma}_{vk}^2)+\sum_{k=1}^{k^B}\hat{E}_{vk}\right), \quad (2)$$

$$\hat{E}_{vk}=-\sum_{l=}^{L_k}\frac{N_{vk1}}{N_v}\log\frac{N_{vk1}}{N_v} \quad (3)$$

حيث ان:-

K^A تمثل العدد الكلى للمتغيرات المستمرة(Continuous Variables) المستخدمة فى عملية العنقدة.

K^B تمثل العدد الكلى للمتغيرات المتقطعة(Categorical Variables) المستخدمة فى عملية العنقدة.

L_K عدد الحالات المتقطعة الموجودة بالنسبة للمتغيرات المتقطعة.

R_K المدى الموجود بالنسبة للمتغيرات المستمرة.

N العدد الكلى لبيانات السجلات .

N_K العدد الكلى لبيانات السجلات داخل العنقود.

$\hat{\sigma}_K^2$ التباين الموجود فى المتغيرات المستمرة من جملة جميع البيانات.

$\hat{\sigma}_{jK}^2$ التباين الموجود فى المتغيرات المستمرة من جملة العنقود(j).

N_{jKl} عدد سجلات البيانات فى العنقود (j) من المتغيرات المتقطعة(Categorical Variables) والتى اخذت من جملة المتغيرات المتقطعة(Categorical Variables).

$d(j,s)$ المسافة بين العنقود (j) والعنقود(s).

$<j,s>$ الفهرس الذى يمثل هيئة العنقود لجمع العنقود (j) والعنقود(s).

وبوضع المتغيرات داخل المعادلات تنتج قيم توضح كيفية قياس المسافة بين العناقيد فى هذه الخوارزمية. ولكن هناك ملحوظة وهى ان هذه هى الطريقة المتبعة عندما تكون المتغيرات من النوع المتقطع(Set data type) اما فى حالة المتغيرات المستمرة(Continuous data type) نستخدم طريقة اخرى تسمى بالـ(Euclidean distance)

وهذه الطريقة تنطبق فقط اذا كانت كل المتغيرات من النوع (Continuous Variables). اما فى حالة المتغيرات تكون من النوعين معا (مستمرة ومتقطعة) نستخدم الطريقة الاولى(Log-likelihood distance)ولكن قيم المتغيرات المستمرة تطبق فى المعادلة بقيمة(0) [8] .

دراسة الحـالــة والنتائج

(Case Study)

1.3 جمع البيانات (Data Collection)

تم جمع البيانات من مستشفى الذرة للطب النووى بالخرطوم حيث كانت البيانات فى شكل جداول حسابية(Microsoft Excel file) وإحتوت على بعض المشاكل مثل القيم المفقودة والقيم الغير منطقية وتمت معالجتها بواسطة واحدة من خطوات اكتشاف المعرفة

هذه البيانات جمعت من نظام حاسوبى خاص بمنظمة الصحة العالمية وكان الغرض من هذا النظام تحليل بيانات السرطان فى فرنسا بأيدى فرنسية فتم أخذ هذه البيانات واجريت عليها عمليات التنقيب باستخدام خوارزمية العنقدة(Two-Step) . البيانات التى جمعت حوت على نوعين من المتغيرات متغيرات الرقمية (Numerical Attribute) ومتغيرات تصنيفية أو مطلقة (Categorical Attribute) والتى تم شرحمها سابقا.

2.3 تجهيز البيانات التى جمعت (Data Set processing)

مثلت البيانات العمود الفقري لعملية التنقيب باستخدام العنقدة أو اكتشاف المعرفة ، والبيانات التي تم التعامل معها لم تكون في الشكل الملائم لعملية التنقيب. فالتحدي الصعب كان هو تجهيزها في الشكل الملائم لعملية التنقيب.

وقد أخذ 70% تقريباً من زمن عملية التنقيب في تجميع وتجهيز البيانات الطبية ، وقد كان التجميع من مصادر مختلفة لذا فان تجهيز البيانات أخذ جهد ووقت حتى تم إخراج النتائج . البيانات التى تم جمعها كانتً غير مكتملة ، ومشوشة ومتناقضة مما قد يخفي أنماط مفيدة ، غير مكتملة بسبب عدم توفر بعض القيم لبعض الخواص والمتغيرات بل كانت هنالك متغيرات غير موجودة أصلاً . أما التشويش فكان بسبب أن البيانات قد أحتوت علي أخطاء و قيم شاذة وكان التناقض في البيانات بسبب التناقض في الرموز و المحتويات فى قاموس البيانات والبيانات التى جمعت نفسها. وتم إختيار البيانات التي لها علاقة مع بعضها البعض : وذلك عن طريق اختيار بعض الخواص وعمل ترشيح لبعضها حيث تم مسح البيانات الشاذة و إزالة السجلات المتكررة . وايضا تم تقليل البيانات: وذلك عن طريق أخذ عينات.

3.3 مشاكل البيانات وكيفية تنقيتها (Data Cleansing)

كانت هنالك مشاكل مختلفة للبيانات فيما يلي بعض هذه المشاكل وكيفية حلها :

1- البيانات الناقصة (Missing Data)

نقول أن البيانات ناقصة إذا كانت بها قيم غير موجودة فى الحقل(Null value) فقد احتوت البيانات التى جمعت على عدد من البيانات المفقوده او الناقصة وتمت معالجتها بوضع قيم جديدة مثل وضح الرقم (10101) للبيانات المفقودة فى حقل مكان الميلاد وعملية معلجة القيم المفقودة إما بحذف السجل او حساب المتوسط للقيم ووضعها مكان القيمة المفقودة. وذلك لان معظم عملية التنقيب تفترض عدم وجود قيم ناقصة.

توجد بيانات مفقودة فى حقل الوظيفة وكان عددها 320 قيمة ولهذا السبب لم تدخل ضمن العناقيد التى نتجت ضمن نموذج العنقدة الذى طبق .

2- البيانات الشاذة او الغير منطقية (outliers data)

هي القيم التي تكون بعيدة المدي الطبيعي للبيانات، أي هي القيم التي لديها انحرافات كثيرة (Standard deviation) من المتوسط (Mean) فقد احتوت البيانات التى جمعت على عدد من البيانات الشاذه مثل بعض هذه القيمة فى حقل عمر المريض (600) وتمت معالجتها بحذف بعضها وايجاد المتوسط لبعضها وذلك لان معظم عملية التنقيب قد تنتج قيم غير صحيحة اذا كانت بها قيم شاذة .

جدول(1.2)– أ وصف البيانات الخام (Data set Description)

خصائص البيانات	متعددة
أنواع البيانات	رقمية ورمزية
المهام الرتبطة	العنقدة
الاهداف	إستخراج معلومات جديدة من مجموعة بيانات مرضى السرطان
عدد السجلات	4250
عدد الحقول	52

هل توجد قيم شاذه	**نعم**
الأدوات المستخدمة	**MS-،oracle9i،Oraloader DTM scrubber،Excel**
المجال	**مرض السرطان**
السنة التى جمعت فيها البيانات	**2007**
القيم المفقودة	**400**
البيانات النظيفة	**4058**

تم أختيار بعض الحقول لعملية التنقيب بواسطة خبير من المستشفى التى جمعت منها البيانات حيث كان الاختيار على البيانات التى من المحتمل ان تنتج معلومات وهى البيانات المهمة لدى العاملين والخبراء بالمستشفى والحقول التى استخدمت فى بناء نموذج العنقدة هى:

1- العمر (Age) من النوع (number) وهى نوع بيانات(Data Type) لتمثيل المتغيرات المستمرة (Continuous Variables) حيث يمثل عمر المريض.

2- النوع (Sex) ويقبل القيمة (1) وتمثل ذكر والقيمة(2) وتمثل أنثى ، من النوع (number)وهى نوع بيانات(Data Type) .

3- أساس التشخيص (Basis) من النوع (number) وهى نوع من انواع البيانات المستخدمة فى أداة التنقيب والتى تمثل المتقيرات المطلقة(Categorical Variables) ويحتوى الحقل على تسعة قيم يمكن ان يقبلها الحقل منها الفحص المعملى، وأخذ عينة من المريض وفحصها معمليا، تشخيص المرض بعد الوفاة، كيفية تشخيص المرض غير واضحة وغيرها من القيم مع العلم بان كل قيمة يقابلها رقم معين.

4- تاريخ التشخيص (Incidence) من النوع (number) ويحتوى الحقل على تاريخ تشخيص المرض.

5- نوع الخلايا المصابة(Morph) من النوع (number) ويحتوى الحقل على عدد من القيم تمثل بأرقام وكل رقم يمثل نوع من انواع الخلايا التى اصابها مرض السرطان.

6- الحالة الأجتماعية(Mstatus) من النوع (number) ويحتوى الحقل على ثلاث قيم (1) وتمثل المتزوجين ، (0) تمثل غير المتزوجين، (9) تمثل العناصر الغير واضحة او لاتوجد بيانات توضح أحدى الحالتين السابقتين.

7- الوظيفة(Occup) من النوع (number) ويحتوى الحقل على قيم تمثل وظيفة المريض.

8- مكان الميلاد(Place_birth) من النوع (number) ويحتوى الحقل على قيم تمثل مكان ميلاد المريض.

9- الجهة(Region) من النوع (number) ويحتوى الحقل على قيم تمثل المنطقة التى ينتمى اليها المريض.

10- درجة انتشار المرض(Seer) من النوع (number) ويحتوى الحقل على قيم تمثل درجة انتشار المرض فى جسم المريض.

11- حالة المريض قبل خروجة من المستشفى(Status) من النوع (number) ويحتوى الحقل على قيم تمثل حالة المريض قبل خروجة من المستشفى .

12-نوع السرطان(Topo) من النوع (number) ويحتوى الحقل على قيم تمثل نوع السرطان الذى كان يعانى منه المريض.

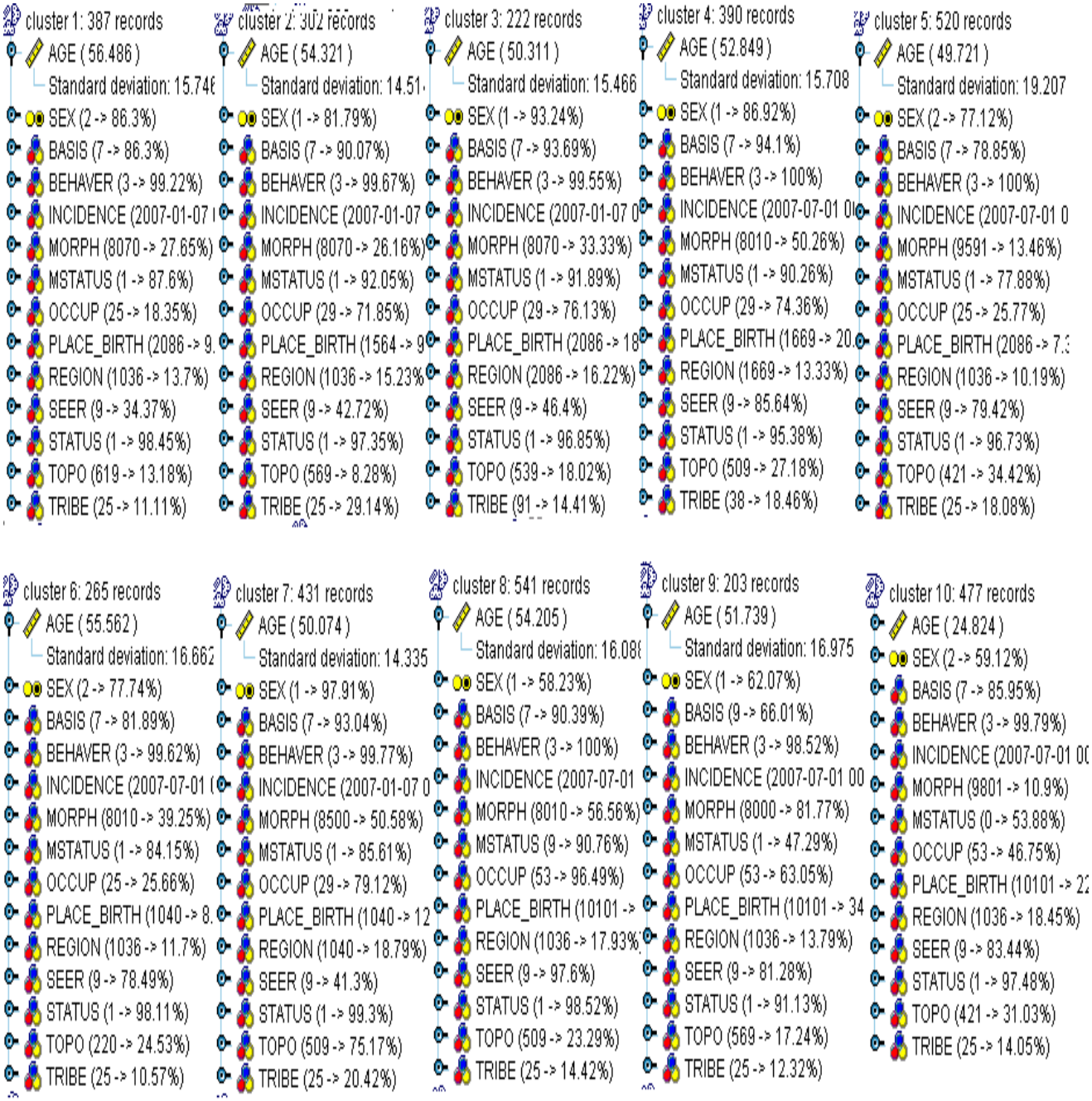
13- القبيلة(Tripe) من النوع (number) ويحتوى الحقل على قيم تمثل نوع القبيلة التى كان ينتمى اليها المريض.

14- سلوك المرض(Behavior) من النوع (number) ويحتوى الحقل على ثلاث قيم تمثل سلوك المرض (1) سرطان حميد، (2) سرطان بين الحميد والخبيث، (3) سرطان خبيث.

4.3 نتائج العنقدة في التنقيب

تم جمع البيانات من مستشفى الذرة للطب النووي وكانت ما يقارب 4250 سجل وأجريت عليها خطوات اكتشاف المعرفة ، ثم أجريت عليها عمليات تحليل البيانات بواسطة التنقيب في البيانات وهو أسلوب غير الأسلوب الاحصائى لتحليل البيانات والتى استخدمت فيها خوارزميات العنقدة وبعد تحليل هذه البيانات نتجت لدينا النتائج التالية:

عندما تم التقسيم الى عشرة عناقيد كما هو موضح فى الشكل 1.3

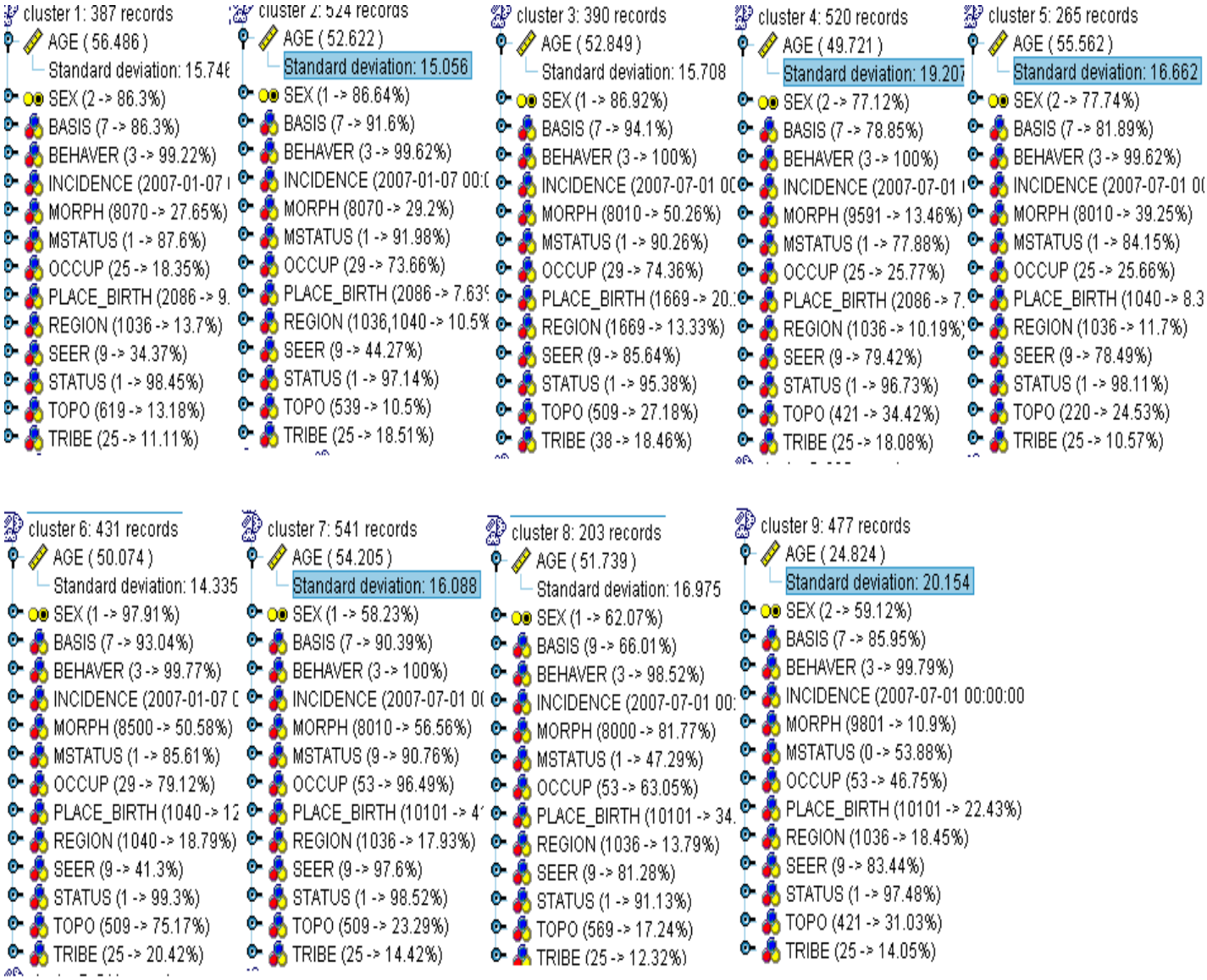


الشكل (1.3) التقسيم الى عشرة عناقيد

يحوى الشكل عشرة أشكال جزئية كل شكل جزئي يمثل عنقود وكل سطر فى الشكل الجزئي يحوى أسم الحقل وبين القوسين يحوى القيمة الغالبة للحقل (أرجع ملحق (أ)- قاموس البيانات) ونسبتها النويه فى العنقود.

الجدول أدناه يوضح نتائج العنقدة للشكل أعلاه

وعند التقسيم الى تسع عناقيد ظهرت نتائج التجارب كما نتجت عند التتقسيم الى عشرة عناقيد ولكن بدمج العنقود الثانى والثالث كما هو موضح فى الشكل 2.3 .

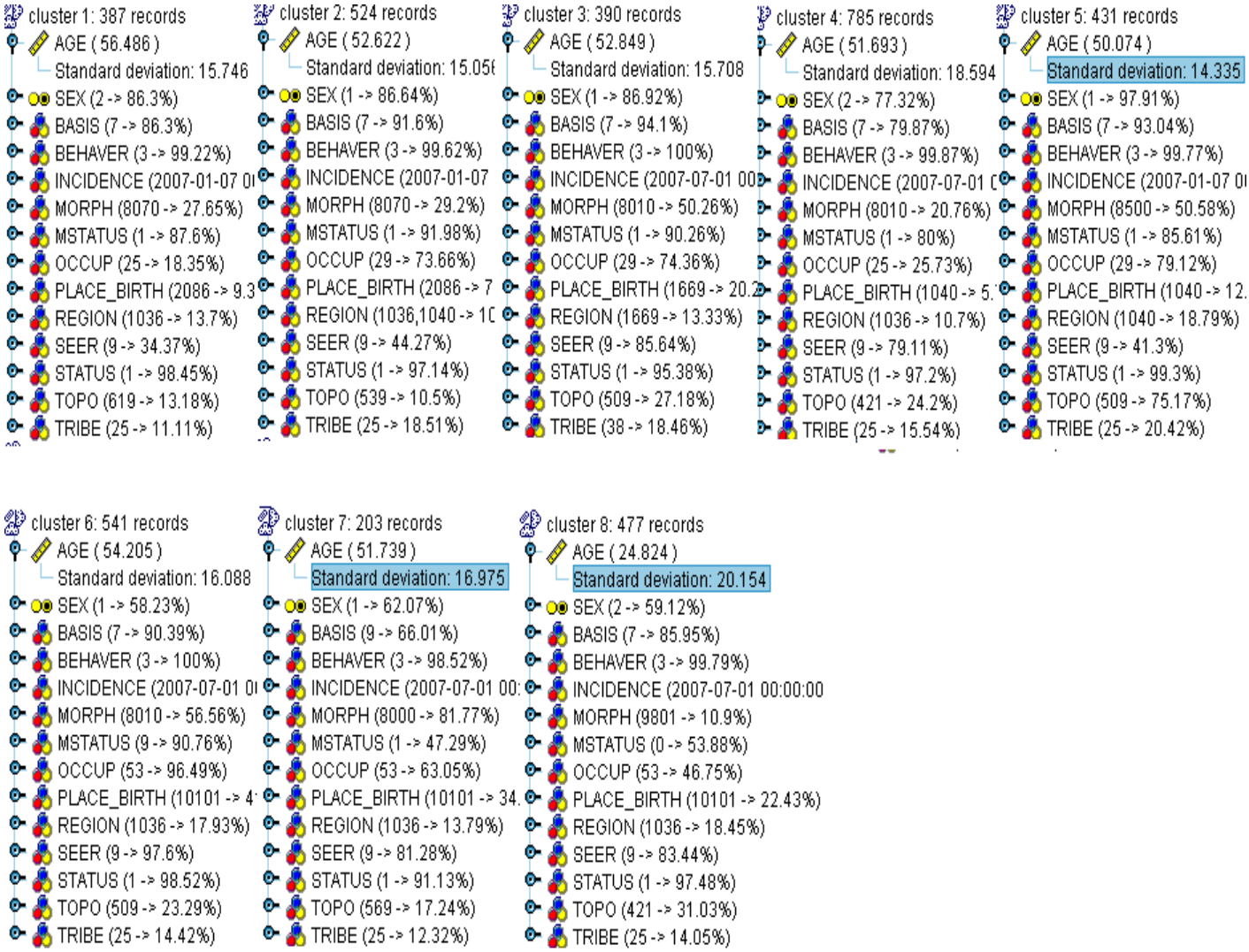


الشكل (2.3) تسعة عناقيد و العنقود الثانى والثالث عند الدمج

وصف العنقودين المدمجين كما يلي:

نلاحظ أن هذه المجموعة تحتوى على 524 حالة، كانت (86%) منها ذكور أعمارهم تتراوح مابين (37-67) سنة وكان تشخيص المرض عن طريق اخذ عينة من المريض وفحصها معمليا والخلايا المصابة بالسرطان هي (Squamous cell carcinoma، NOS) بنسبة 29% من المجموعة وهى اكبر نسبة من بين عناصر المجموعة فى تشخيص نوع خلايا السرطان وتميزت هذه المجموعة بانهم متزوجين بنسبة (91%) وتميزت المجموعة بان عناصرها من النساء كن ربات منزل بينما تميزت ايضا المجموعة اغلبهم كان يقطن مدينة الخرطوم وامدرمان بنسبة (10%)، واتضح أن هذه المجموعة اغلبها مصابة بسرطان البروستات (Meningeal sarcomatosis Cancer) بنسبة (10%).

وعند التقسيم الى ثمانية عناقيد ظهرت نتائج التجارب كما نتجت عند التقسيم الى عشرة وتسعة عناقيد ولكن بدمج العنقود الرابع والخامس كما هو موضح فى الشكل 3.3.

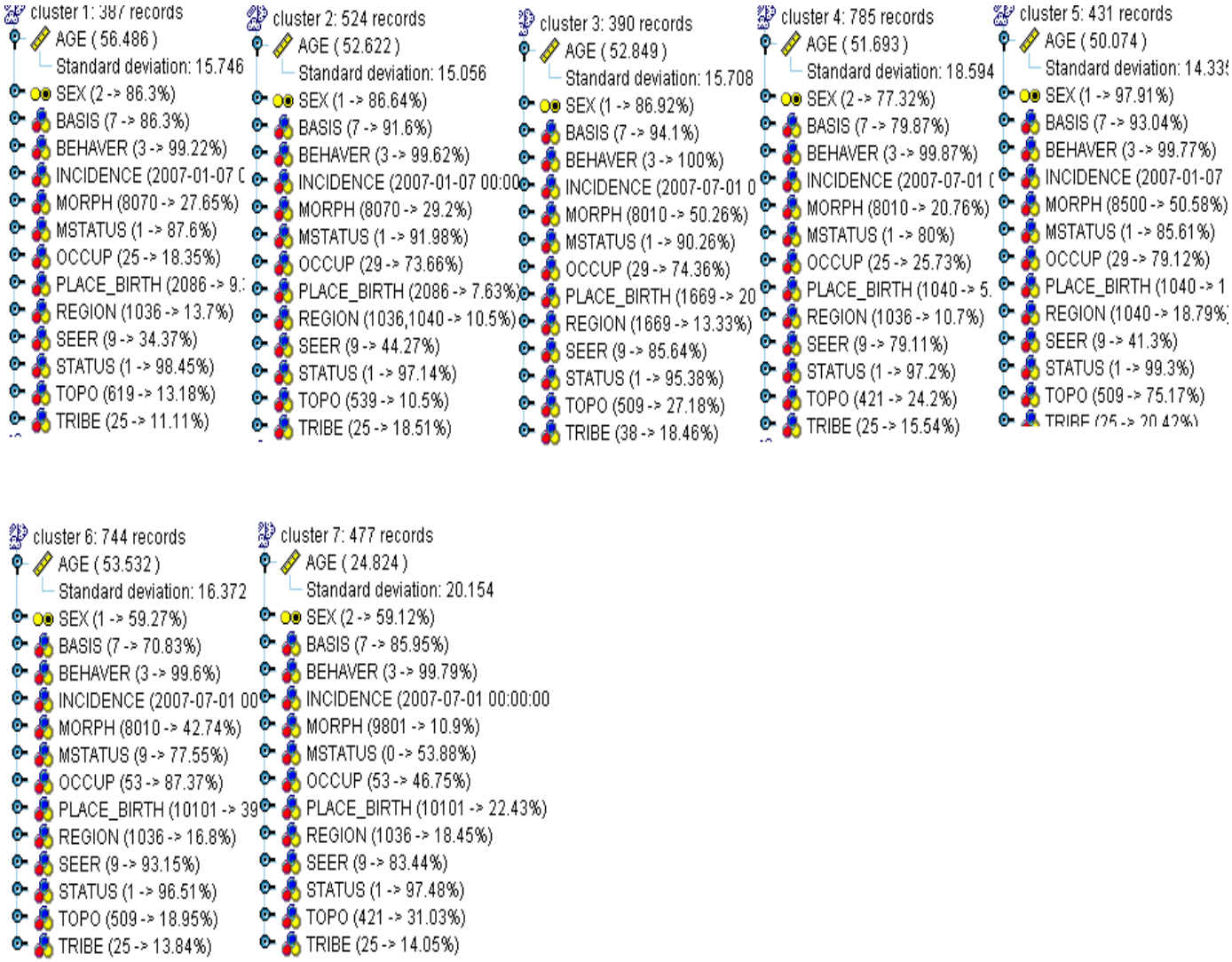


الشكل (3.3) ثمانية عناقيد و العنقود الرابع والخامس عند الدمج

وصف العنقودين المدمجين كما يلي:

نلاحظ أن هذه المجموعة تحتوى على 785 حالة، كانت (77%) منها إناث وأعمارهن تتراوح مابين (33-69) سنة وكان تشخيص المرض عن طريق اخذ عينة من المريض وفحصها معمليا والخلايا المصابة بالسرطان هي (Carcinoma NOS)، بنسبة 20% من المجموعة وهى اكبر نسبة من بين عناصر المجموعة فى تشخيص نوع خلايا السرطان وتميزت هذه المجموعة بانهم متزوجين بنسبة (80%) وتميزت المجموعة بان عناصرها كانت من المزارعين بينما تميزت ايضا المجموعة بان عناصرها مولودون بمدينة امدرمان وغيرها من المدن المختلفة، اغلبهم كان يقطن مدينة الخرطوم بنسبة (10%)، واتضح أن هذه المجموعة اغلبها مصابة بسرطان الدم (Blood Cancer) بنسبة (24%).

وعند التقسيم الى سبعة عناقيد ظهرت نتائج التجارب كما نتجت عند التقسيم الى عشرة وتسعة عناقيد ولكن بدمج العنقود السادس السابع كما هو موضح فى الشكل 4.3.

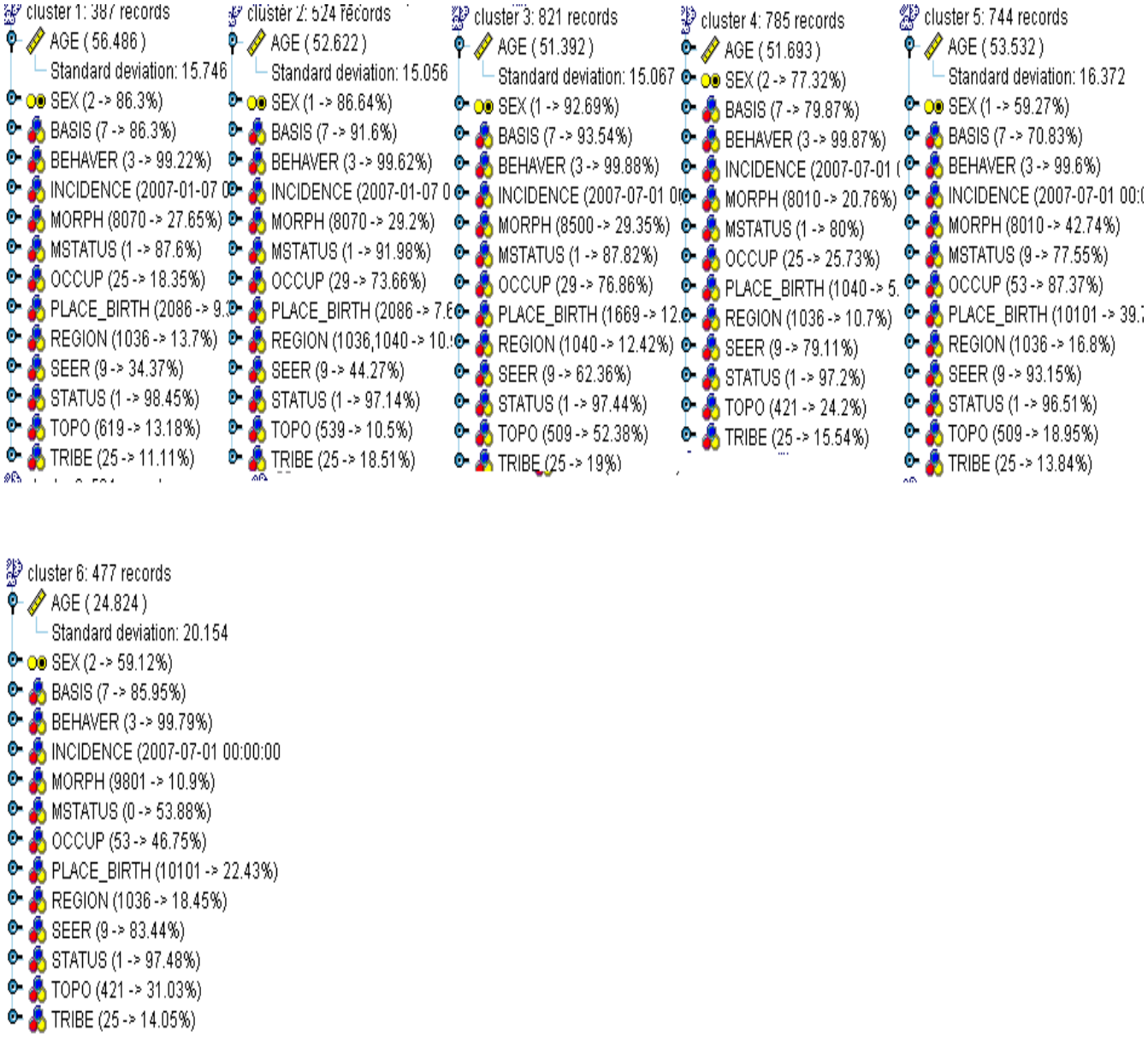


الشكل (4.3) سبعة عناقيد و العنقود السادس والسابع عند الدمج

وصف العنقودين المدمجين كما يلي:

نلاحظ أن هذه المجموعة تحتوى على 744 حالة، كانت (59%) منها ذكور أعمارهم تتراوح مابين (37-69) سنة وكان تشخيص المرض عن طريق اخذ عينة من المريض وفحصها معمليا والخلايا المصابة بالسرطان هي (Carcinoma NOS)، بنسبة 42% من المجموعة وهى اكبر نسبة من بين عناصر المجموعة فى تشخيص نوع خلايا السرطان تميزت بان (77%) لم تكن حالتهم الاجتماعية واضحة وتميزت المجموعة بان عناصرها كانوا من الموظفين بالمعاش بينما تميزت ايضا المجموعة بان عناصرها لم تكن اماكن ولادتهم واضحه، اغلبهم كان يقطن مدينة الخرطوم بنسبة (16%)، واتضح أن هذه المجموعة اغلبها مصابة بسرطان الثدى (Breast Cancer) بنسبة (18%).

وعند التقسيم الى ستة عناقيد ظهرت نتائج التجارب كما نتجت عند التتقسيم الى سبعة عناقيد ولكن بدمج العنقود الثالث والخامس كما هو موضح فى الشكل 5.3.

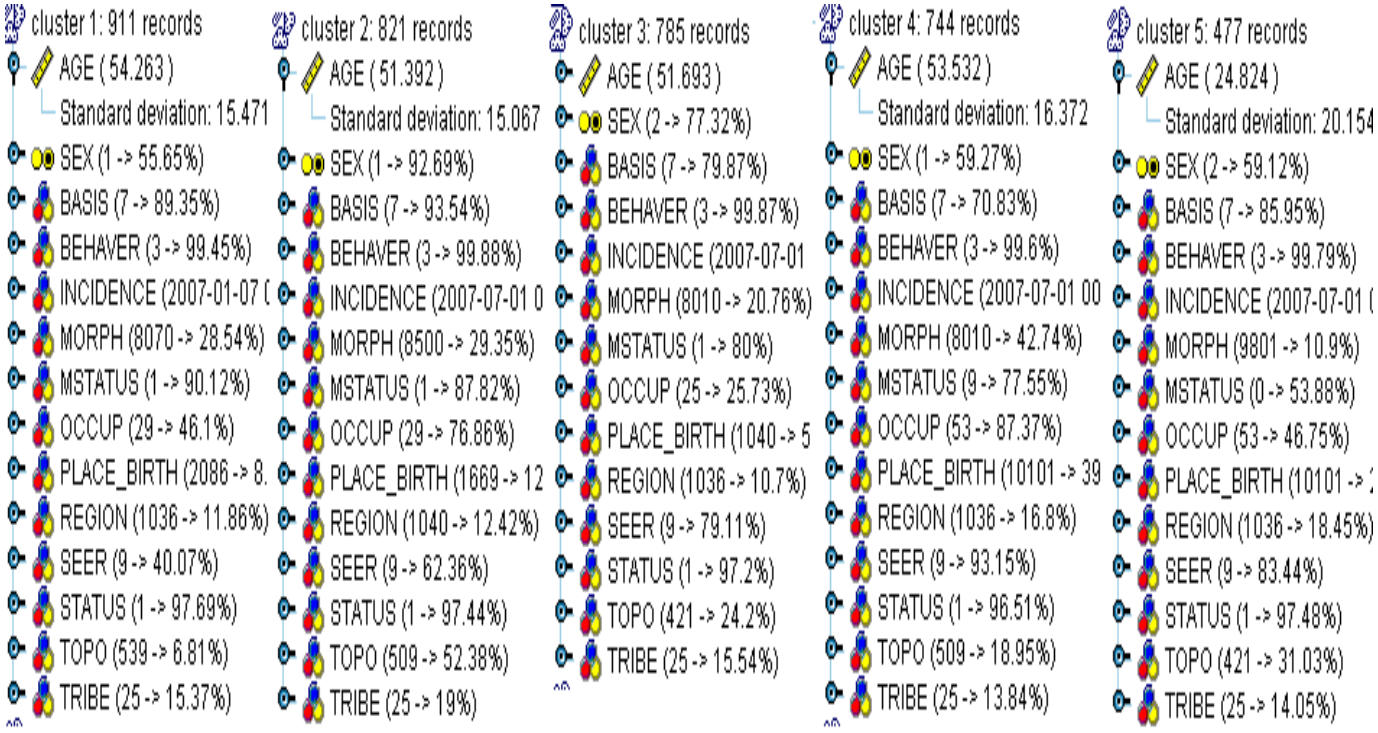


الشكل (5.3) ستة عناقيد و العنقود الثالث والخامس عند الدمج

وصف العنقودين المدمجين كما يلي:

نلاحظ أن هذه المجموعة تحتوى على 821 حالة، كانت (92%) منها ذكور أعمارهم تتراوح مابين (37-69) سنة وكان تشخيص المرض عن طريق اخذ عينة من المريض وفحصها معمليا والخلايا المصابة بالسرطان هي (Infiltrating duct carcinoma NOS)، بنسبة 29% من المجموعة وهى اكبر نسبة من بين عناصر المجموعة فى تشخيص نوع خلايا السرطان وتميزت هذه المجموعة بانهم متزوجين بنسبة (87%) وتميزت المجموعة بان عناصرها من النساء كن ربات منزل بينما تميزت ايضا المجموعة بان عناصرها مولودون بمدينة دنقلا وغيرها من المدن المختلفة، اغلبهم كان يقطن مدينة امدرمان بنسبة (12%)، واتضح أن هذه المجموعة اغلبها مصابة بسرطان الثدى (Breast Cancer) بنسبة (52%).

وعند التقسيم الى خمسة عناقيد ظهرت نتائج التجارب كما نتجت عند التتقسيم الى ستة عناقيد ولكن بدمج العنقود الاول والثانى كما هو موضح فى الشكل 6.3.

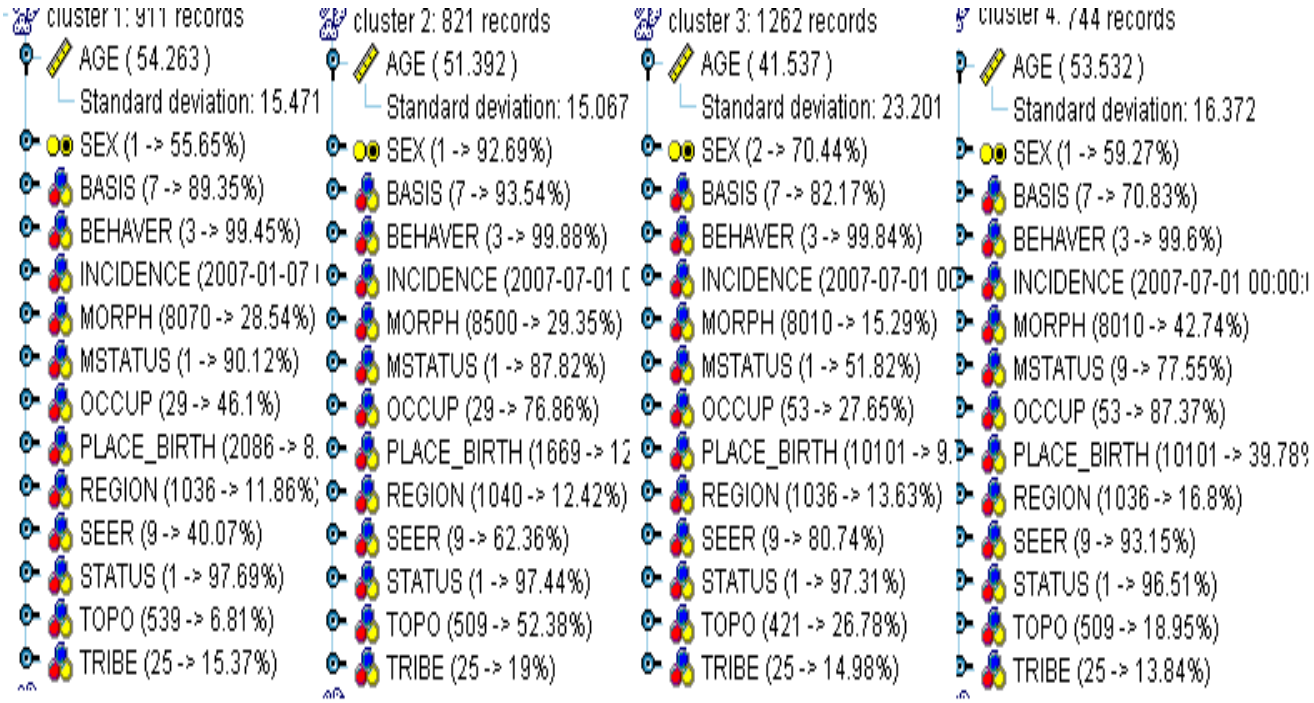


الشكل (6.3) خمسة عناقيد و العنقود الاول والثانى عند الدمج

وصف العنقودين المدمجين كما يلي:

نلاحظ أن هذا المجموعة تحتوى على 911 حالة، كانت (55%) منها ذكور أعمارهم تتراوح مابين (39-69) سنة وكان تشخيص المرض عن طريق اخذ عينة من المريض وفحصها معمليا والخلايا المصابة بالسرطان هي (Squamous cell carcinoma، NOS) بنسبة 28% من المجموعة وهى اكبر نسبة من بين عناصر المجموعة فى تشخيص نوع خلايا السرطان وتميزت هذه المجموعة بانهم متزوجين بنسبة (90%) وتميزت المجموعة بان عناصرها من النساء كن ربات منزل بينما تميزت ايضا المجموعة بان عناصرها مولودون بمدينة شيكان وغيرها من المدن المختلفة، اغلبهم كان يقطن مدينة الخرطوم بنسبة (11%)، واتضح أن النساء من هذه المجموعة اغلبهن أصبن بسرطان عنق الرحم (Cervix uteri Cancer) بنسبة (6%).

وعند التقسيم الى أربعة عناقيد ظهرت نتائج التجارب كما نتجت عند التتقسيم الى خمسة عناقيد ولكن بدمج العنقود الثالث والخامس كما هو موضح فى الشكل 7.3.



الشكل (7.3) أربعة عناقيد و العنقود الثالث والخامس عند الدمج

وصف العنقودين المدمجين كما يلي:-

نلاحظ أن هذه المجموعة تحتوى على 1262 حالة، كانت (70%) منها إناث أعمارهن تتراوح مابين (18-64) سنة وكان تشخيص المرض عن طريق اخذ عينة من المريض وفحصها معمليا والخلايا المصابة بالسرطان هي (Carcinoma NOS)، بنسبة 15% من المجموعة وهى اكبر نسبة من بين عناصر المجموعة فى تشخيص نوع خلايا السرطان وتميزت هذه المجموعة بانهم متزوجين بنسبة (51%) وتميزت المجموعة بان عناصرها كانوا من الموظفين بالمعاش بينما تميزت ايضا المجموعة بان عناصرها لم تكن اماكن ولادتهم واضحه، اغلبهم كان يقطن مدينة الخرطوم بنسبة (13%)، واتضح أن هذه المجموعة اغلبها مصابة بسرطان الدم (Blood Cancer) بنسبة (26%).

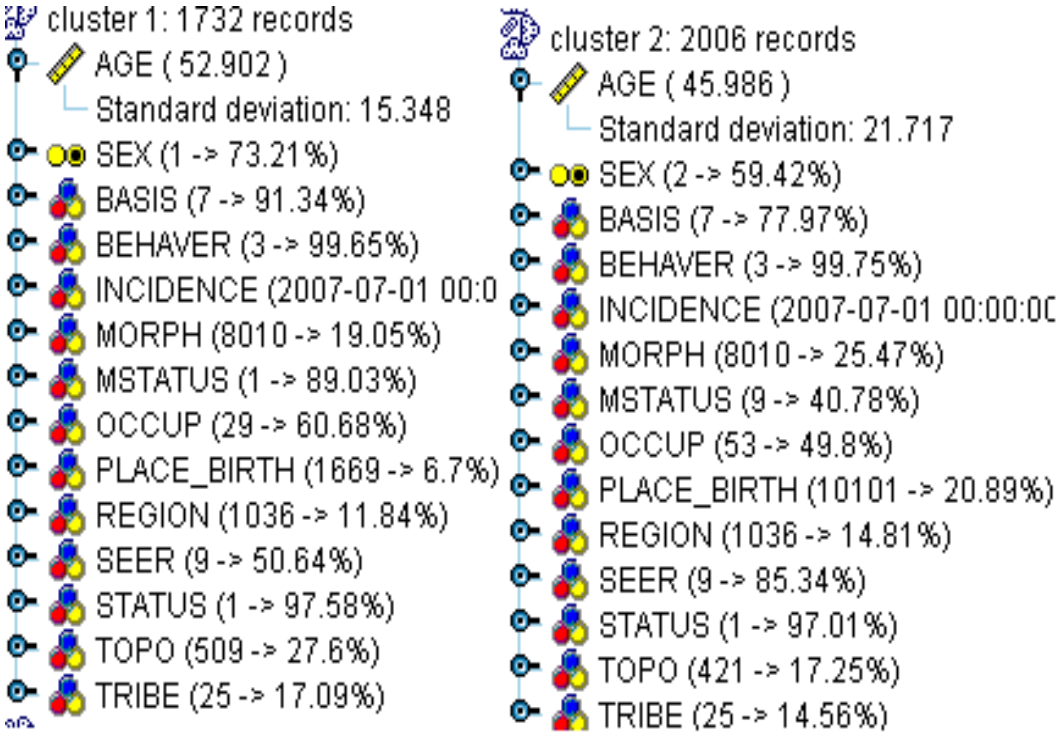
وعند التقسيم الى ثلاثة عناقيد ظهرت نتائج التجارب كما نتجت عند التتقسيم الى أربعة عناقيد ولكن بدمج العنقود الاول والثانى كما هو موضح فى الشكل 8.3.

الشكل (8.3) ثلاثة عناقيد و العنقود الاول والثانى عند الدمج

وصف العنقودين المدمجين كما يلي:

نلاحظ أن هذه المجموعة تحتوى على 1732 حالة، كانت (73%) منها ذكور أعمارهم تتراوح مابين (37-67) سنة وكان تشخيص المرض عن طريق اخذ عينة من المريض وفحصها معمليا والخلايا المصابة بالسرطان هي (Carcinoma NOS)، بنسبة 19% من المجموعة وهى اكبر نسبة من بين عناصر المجموعة فى تشخيص نوع خلايا السرطان وتميزت هذه المجموعة بانهم متزوجين بنسبة (89%) والنساء من هذه المجموعة ربات منزل بنسبة (60%) بينما تميزت ايضا المجموعة بان عناصرها مولودون بمدينة دنقلا وغيرها من المدن المختلفة، اغلبهم كان يقطن مدينة الخرطوم بنسبة (11%)، واتضح أن هذه المجموعة اغلبها مصابة بسرطان الثدى (Breast Cancer) بنسبة (27%).

وعند التقسيم الى عنقودين ظهرت نتائج التجارب بدمج العنقود الثانى والثالث والتى ظهرت عندما قسمت البيانات الى ثلاثة مجموعات كما هو موضح فى الشكل 9.3.



الشكل (9.3) ثلاثة عناقيد و العنقود الثانى والثالث عند الدمج

وصف العنقودين المدمجين كما يلي:

نلاحظ أن هذه المجموعة تحتوى على 2006 حالة، كانت (59%) منها ذكور أعمارهم تتراوح مابين(24-66) سنة وكان تشخيص المرض عن طريق اخذ عينة من المريض وفحصها معمليا والخلايا المصابة بالسرطان هي (Carcinoma ،(NOS بنسبة 25% من المجموعة وهى اكبر نسبة من بين عناصر المجموعة فى تشخيص نوع خلايا السرطان وهذه المجموعة تميزت بان (40%) لم تكن حالتهم الاجتماعية واضحة وتميزت المجموعة بان عناصرها كانوا من الموظفين بالمعاش، بينما تميزت ايضا المجموعة بان عناصرها لم تكن اماكن ولادتهم واضحه او معروفة، واغلبهم كان يقطن مدينة الخرطوم بنسبة (14%)، واتضح أن هذه المجموعة اغلبها مصابة بسرطان الدم (Blood Cancer) بنسبة (17%).

5.3 تقيم ومناقشة النتائج

من تحاليل النتائج أعلاه نجد من الوصف انه لا يوجد عنقود غير مألوف أى لم نتحصل على معلومة جديدة على حسب التجارب التى اجريت وقد يكون هذا بسبب ضعف تبيان البيانات الطبية او لان البيانات تحتاج الى زيادة أكثر من المستخدمة فى خوارزمية التنقيب المطبقة ، يمكن ان نقول ان خوارزمية (Tow-Step) لم تنجح فى ايجاد عناقيد جديدة وانما قامت بعملية وصف للبيانات كما موضح فى النتائج.

الباب الرابع

التوصـــــــــيات

من خلال النتائج التى ظهرت لنا أوصى بالاتى:

1- زيادة كمية البيانات المستخدمة فى عملية التنقيب لمرضى السرطان.

2- استخدام طريقة تنقيب بيانات غير العنقدة(Clustering) مثل التصنيف(Classification) اوغيرها من طرق تنقيب البيانات.

3- توفير بيانات جاهزة لأن خوارزميات التنقيب تفترض أن المدخلات من البيانات موزعة بشكل جيد ولا تحتوي علي قيم مفقودة أو خاطئة وأن جميع الميزات الجيدة متوفره فيها فبالتالى يجب تجهيز البيانات ووضعها فى مستودعات خاصة بالتنقيب و تجهيز البيانات عملية مهمة لذلك يجب علي الباحثون أن يكثفوا الجهود لتطوير تقنيات ملائمة لإدارة واستخدام البيانات بكفاءة ، فبينما نجد أن تقنيات التنقيب في البيانات تدعم تطبيقات تحليل البيانات داخل المنظمات فيجب أن يكون من المحتمل وجود تقنيات لتجهيز بيانات ممتازة من البيانات الخام حتي نستطيع إكتشاف المعرفة بكفاءة وبنوعية ممتازة من البيانات المعطاة، لذلك نجد أن تطوير تقنيات ومنهجيات لتجهيز البيانات تعتبر صعبة وبها قدر من التحدي وعملية مهمة.

4- واخيرا علم تنقيب البيانات من العلوم الجديدة فى مجال الحاسوب والمهمة جدا فى الرؤى المستقبلية وتتطوير المؤسسات والشركات والتى بالتالى تطّور بلادنا وتقودها نحو التقدم فيجب الاهتمام به والنظر اليه بعين فاحصة وتطبيقة على كل المؤسسات التى ينجح فيها.

الخـــــاتمــة

الحمد لله الذى وفقنى لإكمال هذا البحث حتى يكون لبنة يقتدى بها كل باحث فى مجال تنقيب البيانات ومن الضروري لبحث حل مشكلة ما أن يتم تحديدها بشكل دقيق، وجمع كافة المعلومات المتعلقة بها، وتبدأ هذه العملية بمراجعة وافية لطبيعة عمل المؤسسة المعنية وأهدافها والبدء بتحديد طبيعة المشكلة وأسبابها وجذورها وتفرعاتها، وليس بالضرورة أن تكون هناك مشكلة بمعنى الكلمة ولكن قد تكون المشكلة مصاغة على شكل هدف معين أو وصف محدد. وتعتبر من ضمن العمليات المهمة في مراحل التحضير، حيث تم بناء قاعدة بيانات خاصة بعملية التنقيب والتى شملت على تعديلات كثيرة حتى ثم البدء باستكشاف البيانات وتحديد الحقول الأكثر أهمية والقيم التي يمكن من خلالها تقييم أهداف التحليل والتنقيب، .كما تم تحديد المتغيرات والسجلات وتخليق متغيرات جديدة

بحسب الحاجة.وايضا تم إختيار أدوات التحليل الإحصائي و خوارزمية التنقيب المناسبة (Tow-Step) وذلك وفق العديد من الاعتبارات التي يمكن الاستناد عليها، كالغرض من الخوارزمية وإمكانية تطبيقها وتحقيقها لهذا الغرض. وبعدها تم تتطبيق نموذج التحليل أو التنقيب وإجراء كافة التعديلات المناسبة مع مراعاة أهداف المشكلة وبحث إمكانية حلها بالنموذج الذي تم بناءه، وأخيراً تم استخراج النتائج وتفسيرها وتوضيحها على حسب النتائج التي تم التوصل إليها واخيرا كانت هنالك بعض التوصيات التى خرج بها البحث حيث تكون أساس عمل للباحثين من بعدى.

References المراجع

1\http://ar.wikipedia.org/wiki/%D8%AA%D9%86%D9%82%D9%8A%D8%A8_%D9%81%D9%8A_%D8%A7%D9%84%D8%A8%D9%8A%D8%A7%D9%86%D8%A7%D8%AA . Access date 9/10/2008 03/05/22

2\ http://informatics.gov.sa/ access date 9/10/2008 01-11-1

3 \ A.K. JAIN ،Michigan State University، and M.N. MURTY، Indian Institute of Science ،and P.J. FLYNN Ohio State University ،" Data Clustering " Ohio State University، ACM Computing Surveys، Vol. 31، No. 3، September 1999

4\ Botta، Marco "Clustering Techniques "،Dipartimento di Informatica Universitàdi Torino،www.di.unito.it/~botta/didattica /clustering.html،2003.

5/ الياسري ، اروى عيسى ،هديل شوكت العبيدى . تجربة تصميم مكنز الي باستخدام اساليب استخراج البيانات " Data Mining. _ "بحث غير منشورالقي في مؤتمر بلدية دبي الدولي الثالث للتوثيق والارشفة الالكترونية ايلول 2005.

6 / الياسري ،اروى عيسى ،هديل شوكت العبيدي . التصنيف الالي لمصادر المكتبة باستخدام تقنيات التصنيف Classification Techniques في استخراج البيانات Data Mining بحث غيرمنشور

7\ Agrawal R.‹ Imielinski T.‹ and Swami S.‹ Mining Association rules between sets of items in large databases‹ Proc. of the ACM SIGMOD Conference on Management of Data‹ Washington‹ DC‹ May 1993.

8\ Zhang‹ T.‹ Ramakrishnon‹ R.‹ and Livny M. (1996). BIRCH: An Efficient Data Clustering Method for Very Large Databases. Proceedings of the ACM SIGMOD Conoference on Management of Data‹ p. 103-114‹ Montreal‹ Canada.

Chiu‹ T.‹ Fang‹ D.‹ Chen‹ J.‹ Wang‹ Y.‹ and Jeris‹ C. (2001). A Robust and Scalable

Clustering Algorithm for Mixed Type Attributes in Large Database Environment.

Proceedings of the seventh ACM SIGKDD international conference on knowledge

263.‹discovery and data mining

ملحق (أ) قاموس البيانات

>---------- 1#Record status
0Pending
1Confirmed
2Deleted

>---------- 2#Check status
0Not done
1Done: OK
2Done: Rare
3Done: Invalid

>---------- 3#Person search
0Not done
1Done: OK
2Done: Mult. Prim.
3Done: Duplicate

>---------- 4#Topography
00LIP
01TONGUE - BASE
02TONGUE - OTHER
03GUM
04MOUTH - FLOOR
05MOUTH - PALATE
06MOUTH - OTHER
07MOUTH - PAROTID
08MOUTH - GLANDS
09TONSILS
10OROPHARYNX
11NASOPHARYNX
12PYRIFORM SINUS
13HYPOPHARYNX
14OTHER PHARYNX
15ESOPHAGUS
16STOMACH
17SMALL INTESTINE
18COLON
19RECTOSIGMOID JTN
20RECTUM
21ANUS, ANAL CANAL
22LIVER
23GALL BLADDER
24BILE TRACT
25PANCREAS
26GASTROINTESTINAL
30EAR/NOSE
31ACCESSORY SINUSES
32LARYNX

33TRACHEA
34BRONCHUS, LUNG
37THYMUS
38HEART, PLEURA
39RESPIRATORY TRACT
40BONES - LIMBS
41BONES - OTHER
42BLOOD, SPLEEN
44SKIN
47PERIPHERAL NERVES
48PERITONEUM
49SOFT TISSUES
50BREAST
51VULVA
52VAGINA
53CERVIX UTERI
54CORPUS UTERI
55UTERUS, NOS
56OVARY
57FEMALE GEN.- OTHER
58PLACENTA
60PENIS
61PROSTATE
62TESTIS
63MALE GEN.- OTHER
64KIDNEY
65RENAL PELVIS
66URETER
67BLADDER
68URETHRA, OTHER
69EYE
70MENINGES
71BRAIN
72NERVOUS SYSTEM
73THYROID
74ADRENAL GLAND
75ENDOCRINE GLANDS
76ILL-DEFINED
77LYMPH NODES
80UNKNOWN
000C00.0 External upper lip
001C00.1 External lower lip
002C00.2 External lip, NOS
003C00.3 Mucosa of upper lip
004C00.4 Mucosa of lower lip
005C00.5 Mucosa of lip, NOS
006C00.6 Commissure of lip
008C00.8 Overl. lesion of lip
009C00.9 Lip, NOS
019C01.9 Base of tongue, NOS

020C02.0 Dorsal surface of tongue, NOS
021C02.1 Border of tongue
022C02.2 Ventral surface of tongue, NOS
023C02.3 Anterior 2/3 of tongue, NOS
024C02.4 Lingual tonsil
028C02.8 Overl. lesion of tongue
029C02.9 Tongue, NOS
030C03.0 Upper gum
031C03.1 Lower gum
039C03.9 Gum, NOS
040C04.0 Anterior floor of mouth
041C04.1 Lateral floor of mouth
048C04.8 Overl. lesion of floor of mouth
049C04.9 Floor of mouth, NOS
050C05.0 Hard palate
051C05.1 Soft palate, NOS
052C05.2 Uvula
058C05.8 Overl. lesion of palate
059C05.9 Palate, NOS
060C06.0 Cheek mucosa
061C06.1 Vestibule of mouth
062C06.2 Retromolar area
068C06.8 Overl. lesion of other/unspec. parts of mo
069C06.9 Mouth, NOS
079C07.9 Parotid gland
080C08.0 Submandibular gland
081C08.1 Sublingual gland
088C08.8 Overl. lesion of major salivary gland
089C08.9 Major salivary gland, NOS
090C09.0 Tonsillar fossa
091C09.1 Tonsillar pillar
098C09.8 Overl. lesion of tonsil
099C09.9 Tonsil, NOS
100C10.0 Vallecula
101C10.1 Anterior surface of epiglottis
102C10.2 Lateral wall of oropharynx
103C10.3 Posterior wall of oropharynx
104C10.4 Branchial cleft
108C10.8 Overl. lesion of oropharynx
109C10.9 Oropharynx, NOS
110C11.0 Superior wall of nasopharynx
111C11.1 Posterior wall of nasopharynx
112C11.2 Lateral wall of nasopharynx
113C11.3 Anterior wall of nasopharynx
118C11.8 Overl. lesion of nasopharynx
119C11.9 Nasopharynx, NOS
129C12.9 Pyriform sinus
130C13.0 Postcricoid region
131C13.1 Aryepiglottic fold
132C13.2 Posterior wall of hypopharynx

138C13.8 Overl. lesion of hypopharynx
139C13.9 Hypopharynx, NOS
140C14.0 Pharynx, NOS
141C14.1 Laryngopharynx
142C14.2 Waldeyer's ring, NOS
148C14.8 Overl. lesion of lip, oral cavity, pharynx
150C15.0 Cervical esophagus
151C15.1 Thoracic esophagus
152C15.2 Abdominal esophagus
153C15.3 Upper third of esophagus
154C15.4 Middle third of esophagus
155C15.5 Lower third of esophagus
158C15.8 Overl. lesion of esophagus
159C15.9 Oesophagus, NOS
160C16.0 Cardia, NOS
161C16.1 Fundus of stomach
162C16.2 Body of stomach
163C16.3 Gastric antrum
164C16.4 Pylorus
165C16.5 Lesser curvature of stomach, NOS
166C16.6 Greater curvature of stomach, NOS
168C16.8 Overl. lesion of stomach
169C16.9 Stomach, NOS
170C17.0 Duodenum
171C17.1 Jejunum
172C17.2 Ileum
173C17.3 Meckel's diverticulum
178C17.8 Overl. lesion of small intestine
179C17.9 Small intestine
180C18.0 Cecum
181C18.1 Appendix
182C18.2 Ascending colon
183C18.3 Hepatic flexure of colon
184C18.4 Transverse colon
185C18.5 Splenic flexure of colon
186C18.6 Descending colon
187C18.7 Sigmoid colon
188C18.8 Overl. lesion of colon
189C18.9 Colon, NOS
199C19.9 Rectosigmoid junction
209C20.9 Rectum, NOS
210C21.0 Anus, NOS
211C21.1 Anal canal
212C21.2 Cloacogenic zone
218C21.8 Overlap rectum, anus, anal canal
220C22.0 Liver
221C22.1 Intrahepatic bile duct
239C23.9 Gallbladder
240C24.0 Extrahepatic bile duct
241C24.1 Ampulla of Vater

248C24.8 Overl. lesion of biliary tract
249C24.9 Biliary tract, NOS
250C25.0 Head of pancreas
251C25.1 Body of pancreas
252C25.2 Tail of pancreas
253C25.3 Pancreatic duct
254C25.4 Islets of Langerhans
257C25.7 Other specified parts of pancreas
258C25.8 Overl. lesion of pancreas
259C25.9 Pancreas, NOS
260C26.0 Intestinal tract, NOS
268C26.8 Overl. lesion of digestive system
269C26.9 Gastrointestinal tract, NOS
300C30.0 Nasal cavity
301C30.1 Middle ear
310C31.0 Maxillary sinus
311C31.1 Ethmoid sinus
312C31.2 Frontal sinus
313C31.3 Sphenoid sinus
318C31.8 Overl. lesion of accessory sinuses
319C31.9 Accessory sinus, NOS
320C32.0 Glottis
321C32.1 Supraglottis
322C32.2 Subglottis
323C32.3 Laryngeal cartilage
328C32.8 Overl. lesion of larynx
329C32.9 Larynx, NOS
339C33.9 Trachea
340C34.0 Main bronchus
341C34.1 Upper lobe, lung
342C34.2 Middle lobe, lung
343C34.3 Lower lobe, lung
348C34.8 Overl. lesion of lung
349C34.9 Lung, NOS
379C37.9 Thymus
380C38.0 Heart
381C38.1 Anterior mediastinum
382C38.2 Posterior mediastinum
383C38.3 Mediastinum, NOS
384C38.4 Pleura, NOS
388C38.8 Overl. lesion of heart, mediastinum, pleur
390C39.0 Upper respiratory tract
398C39.8 Overl. lesion of respiratory system
399C39.9 Ill-defined sites within respiratory syste
400C40.0 Long bones of upper limb, scapula
401C40.1 Short bones of upper limb
402C40.2 Long bones of lower limb
403C40.3 Short bones of lower limb
408C40.8 Overl. lesion of bones of limb
409C40.9 Bone of limb, NOS

410C41.0 Bones of skull and face
411C41.1 Mandible
412C41.2 Vertebral column
413C41.3 Rib, Sternum, Clavicle
414C41.4 Pelvic bones, Sacrum, Coccyx
418C41.8 Overl. lesion of bones
419C41.9 Bone, NOS
420C42.0 Blood
421C42.1 Bone marrow
422C42.2 Spleen
423C42.3 Reticuloendothelial system, NOS
424C42.4 Hematopoietic system, NOS
440C44.0 Skin of lip, NOS
441C44.1 Eyelid
442C44.2 External ear
443C44.3 Skin, other & unspec parts of face
444C44.4 Skin of scalp and neck
445C44.5 Skin of trunk
446C44.6 Skin of upper limb and shoulder
447C44.7 Skin of lower limb and hip
448C44.8 Overl. lesion of skin
449C44.9 Skin, NOS
470C47.0 Per. nerves & A.N.S. of head, face, neck
471C47.1 Per. nerves & A.N.S. of upper limb, should
472C47.2 Per. nerves & A.N.S. of lower limb, hip
473C47.3 Per. nerves & A.N.S. of thorax
474C47.4 Per. nerves & A.N.S. of abdomen
475C47.5 Per. nerves & A.N.S. of pelvis
476C47.6 Per. nerves & A.N.S. of trunk
478C47.8 Overl. lesion of peripheral nerves & ANS
479C47.9 Autonomic nervous system, NOS
480C48.0 Retroperitoneum
481C48.1 Specified parts of peritoneum
482C48.2 Peritoneum, NOS
488C48.8 Overl. lesion of retroperitoneum & periton
490C49.0 Soft tissues of head, face, & neck
491C49.1 Soft tissues of upper limb, shoulder
492C49.2 Soft tissues of lower limb and hip
493C49.3 Soft tissues of thorax
494C49.4 Soft tissues of abdomen
495C49.5 Soft tissues of pelvis
496C49.6 Soft tissues of trunk
498C49.8 Overl. lesion of soft tissues
499C49.9 Other soft tissues
500C50.0 Nipple
501C50.1 Central portion of breast
502C50.2 Upper-inner quadrant of breast
503C50.3 Lower-inner quadrant of breast
504C50.4 Upper-outer quadrant of breast
505C50.5 Lower-outer quadrant of breast

506C50.6 Axillary tail of breast
508C50.8 Overl. lesion of breast
509C50.9 Breast, NOS
510C51.0 Labium majus
511C51.1 Labium minus
512C51.2 Clitoris
518C51.8 Overl. lesion of vulva
519C51.9 Vulva, NOS
529C52.9 Vagina, NOS
530C53.0 Endocervix
531C53.1 Exocervix
538C53.8 Overl. lesion of cervix uteri
539C53.9 Cervix uteri
540C54.0 Isthmus uteri
541C54.1 Endometrium
542C54.2 Myometrium
543C54.3 Fundus uteri
548C54.8 Overl. lesion of corpus uteri
549C54.9 Corpus uteri
559C55.9 Uterus, NOS
569C56.9 Ovary
570C57.0 Fallopian tube
571C57.1 Broad ligament
572C57.2 Round ligament
573C57.3 Parametrium
574C57.4 Uterine adnexa
577C57.7 Other parts of female genital organs
578C57.8 Overl. lesion of female genital organs
579C57.9 Female genital tract, NOS
589C58.9 Placenta
600C60.0 Prepuce
601C60.1 Glans penis
602C60.2 Body of penis
608C60.8 Overl. lesion of penis
609C60.9 Penis, NOS
619C61.9 Prostate gland
620C62.0 Undescended testis
621C62.1 Descended testis
629C62.9 Testis, NOS
630C63.0 Epididymis
631C63.1 Spermatic cord
632C63.2 Scrotum, NOS
637C63.7 Other parts of male genital organs
638C63.8 Overl. lesion of male genital organs
639C63.9 Male genital organs, NOS
649C64.9 Kidney, NOS
659C65.9 Renal pelvis
669C66.9 Ureter
670C67.0 Trigone of urinary bladder
671C67.1 Dome of urinary bladder

672C67.2 Lateral wall of urinary bladder
673C67.3 Anterior wall of urinary bladder
674C67.4 Posterior wall of urinary bladder
675C67.5 Bladder neck
676C67.6 Ureteric orifice
677C67.7 Urachus
678C67.8 Overl. lesion of bladder
679C67.9 Urinary bladder, NOS
680C68.0 Urethra
681C68.1 Paraurethral gland
688C68.8 Overl. lesion of urinary organs
689C68.9 Urinary system, NOS
690C69.0 Conjunctiva
691C69.1 Cornea, NOS
692C69.2 Retina
693C69.3 Choroid
694C69.4 Ciliary body
695C69.5 Lacrimal gland, NOS
696C69.6 Orbit, NOS
698C69.8 Overl. lesion of eye, adnexa
699C69.9 Eye, NOS
700C70.0 Cerebral meninges
701C70.1 Spinal meninges
709C70.9 Meninges, NOS
710C71.0 Cerebrum
711C71.1 Frontal lobe
712C71.2 Temporal lobe
713C71.3 Parietal lobe
714C71.4 Occipital lobe
715C71.5 Ventricle, NOS
716C71.6 Cerebellum, NOS
717C71.7 Brain stem
718C71.8 Overl. lesion of brain
719C71.9 Brain, NOS
720C72.0 Spinal cord
721C72.1 Cauda equina
722C72.2 Olfactory nerve
723C72.3 Optic nerve
724C72.4 Acoustic nerve
725C72.5 Cranial nerve
728C72.8 Overl. lesion of brain and CNS
729C72.9 Nervous system, NOS
739C73.9 Thyroid gland
740C74.0 Cortex of adrenal gland
741C74.1 Medulla of adrenal gland
749C74.9 Adrenal gland, NOS
750C75.0 Parathyroid gland
751C75.1 Pituitary gland
752C75.2 Craniopharyngeal duct
753C75.3 Pineal gland

754C75.4 Carotid body
755C75.5 Aortic body and other paraganglia
758C75.8 Overl. lesion of endocrine glands
759C75.9 Endocrine gland, NOS
760C76.0 Head, face or neck, NOS
761C76.1 Thorax, NOS
762C76.2 Abdomen, NOS
763C76.3 Pelvis, NOS
764C76.4 Upper limb, NOS
765C76.5 Lower limb, NOS
767C76.7 Other ill-defined sites
768C76.8 Overl. lesion of ill-defined sites
770C77.0 Lymph nodes of head, face and neck
771C77.1 Intrathoracic lymph nodes
772C77.2 Intra-abdominal lymph nodes
773C77.3 Lymph nodes of axilla or arm
774C77.4 Lymph nodes, inguinal region or leg
775C77.5 Pelvic lymph nodes
778C77.8 Lymph nodes of multiple regions
779C77.9 Lymph node, NOS
809C80.9 Unknown primary site

>---------- 5#Morphology
8000Neoplasm, malignant
8001Tumor cells, malignant
8002Malignant tumor, small cell type
8003Malignant tumor, giant cell type
8004Malignant tumor, spindle cell type
8005Malignant tumor, clear cell type
8010Carcinoma, NOS
8011Epithelioma, malignant
8012Large cell carcinoma, NOS
8013Large cell neuroendocrine carcinoma
8014Large cell carcinoma with rhabdoid phenotype
8015Glassy cell carcinoma
8020Carcinoma, undifferentiated, NOS
8021Carcinoma, anaplastic, NOS
8022Pleomorphic carcinoma
8030Giant cell and spindle cell carcinoma
8031Giant cell carcinoma
8032Spindle cell carcinoma, NOS
8033Pseudosarcomatous carcinoma
8034Polygonal cell carcinoma
8035Carcinoma with osteoclast-like giant cells
8041Small cell carcinoma, NOS
8042Oat cell carcinoma
8043Small cell carcinoma, fusiform cell
8044Small cell carcinoma, intermediate cell
8045Combined small cell carcinoma
8046Non-small cell carcinoma

8050Papillary carcinoma, NOS
8051Verrucous carcinoma, NOS
8052Papillary squamous cell carcinoma
8070Squamous cell carcinoma, NOS
8071Squamous cell carcinoma, keratinizing, NOS
8072Squamous cell carcinoma, large cell, nonkeratinizing, NOS
8073Squamous cell carcinoma, small cell, nonkeratinizing
8074Squamous cell carcinoma, spindle cell
8075Squamous cell carcinoma, adenoid
8076Squamous cell carcinoma, microinvasive
8078Squamous cell carcinoma with horn formation
8082Lymphoepithelial carcinoma
8083Basaloid squamous cell carcinoma
8084Squamous cell carcinoma, clear cell type
8090Basal cell carcinoma, NOS
8091Multifocal superficial basal cell carcinoma
8092Infiltrating basal cell carcinoma, NOS
8093Basal cell carcinoma, fibroepithelial
8094Basosquamous carcinoma
8095Metatypical carcinoma
8097Basal cell carcinoma, nodular
8098Adenoid basal carcinoma
8102Trichilemmocarcinoma
8110Pilomatrix carcinoma
8120Transitional cell carcinoma, NOS
8121Schneiderian carcinoma
8122Transitional cell carcinoma, spindle cell
8123Basaloid carcinoma
8124Cloacogenic carcinoma
8130Papillary transitional cell carcinoma
8131Transitional cell carcinoma, micropapillary
8140Adenocarcinoma, NOS
8141Scirrhous adenocarcinoma
8142Linitis plastica
8143Superficial spreading adenocarcinoma
8144Adenocarcinoma, intestinal type
8145Carcinoma, diffuse type
8147Basal cell adenocarcinoma
8150Islet cell carcinoma
8151Insulinoma, malignant
8152Glucagonoma, malignant
8153Gastrinoma, malignant
8154Mixed islet cell and exocrine adenocarcinoma
8155Vipoma, malignant
8156Somatostatinoma, malignant
8157Enteroglucagonoma, malignant
8160Cholangiocarcinoma
8161Bile duct cystadenocarcinoma
8162Klatskin tumor
8170Hepatocellular carcinoma, NOS

8171Hepatocellular carcinoma, fibrolamellar
8172Hepatocellular carcinoma, scirrhous
8173Hepatocellular carcinoma, spindle cell variant
8174Hepatocellular carcinoma, clear cell type
8175Hepatocellular carcinoma, pleomorphic type
8180Combined hepatocellular carcinoma and cholangiocarcinoma
8190Trabecular adenocarcinoma
8200Adenoid cystic carcinoma
8201Cribriform carcinoma, NOS
8210Adenocarcinoma in adenomatous polyp
8211Tubular adenocarcinoma
8214Parietal cell carcinoma
8215Adenocarcinoma of anal glands
8220Adenocarcinoma in adenomatous polyposis coli
8221Adenocarcinoma in multiple adenomatous polyps
8230Solid carcinoma, NOS
8231Carcinoma simplex
8240Carcinoid tumor, NOS
8241Enterochromaffin cell carcinoid
8242Enterochromaffin-like cell tumor, malignant
8243Goblet cell carcinoid
8244Composite carcinoid
8245Adenocarcinoid tumor
8246Neuroendocrine carcinoma, NOS
8247Merkel cell carcinoma
8249A typical carcinoid tumor
8250Bronchiolo-alveolar adenocarcinoma, NOS
8251Alveolar adenocarcinoma
8252Bronchiolo-alveolar carcinoma, non- mucinous
8253Bronchiolo-alveolar carcinoma, mucinous
8254Bronchiolo-alveolar carcinoma, mixed mucinous / non-mucinous
8255Adenocarcinoma with mixed subtypes
8260Papillary adenocarcinoma, NOS
8261Adenocarcinoma in villous adenoma
8262Villous adenocarcinoma
8263Adenocarcinoma in tubulovillous adenoma
8270Chromophobe carcinoma
8272Pituitary carcinoma, NOS
8280Acidophil carcinoma
8281Mixed acidophil-basophil carcinoma
8290Oxyphilic adenocarcinoma
8300Basophil carcinoma
8310Clear cell adenocarcinoma, NOS
8312Renal cell carcinoma, NOS
8313Clear cell adenocarcinofibroma
8314Lipid-rich carcinoma
8315Glycogen-rich carcinoma
8316Cyst-associated renal cell carcinoma
8317Renal cell carcinoma, chromophobe type
8318Renal cell carcinoma, sarcomatoid

8319Collecting duct carcinoma
8320Granular cell carcinoma
8322Water-clear cell adenocarcinoma
8323Mixed cell adenocarcinoma
8330Follicular adenocarcinoma, NOS
8331Follicular adenocarcinoma, well differentiated
8332Follicular adenocarcinoma, trabecular
8333Fetal adenocarcinoma
8335Follicular carcinoma, minimally invasive
8337Insular carcinoma
8340Papillary carcinoma, follicular variant
8341Papillary microcarcinoma
8342Papillary carcinoma, oxyphilic cell
8343Papillary carcinoma, encapsulated
8344Papillary carcinoma, columnar cell
8345Medullary carcinoma with amyloid stroma
8346Mixed medullary-follicular carcinoma
8347Mixed medullary-papillary carcinoma
8350Nonencapsulated sclerosing carcinoma
8370Adrenal cortical carcinoma
8380Endometrioid adenocarcinoma, NOS
8381Endometrioid adenofibroma, malignant
8382Endometrioid adenocarcinoma, secretory variant
8383Endometrioid adenocarcinoma, ciliated cell variant
8384Adenocarcinoma, endocervical type
8390Skin appendage carcinoma
8400Sweat gland adenocarcinoma
8401Apocrine adenocarcinoma
8402Nodular hidradenoma, malignant
8403Malignant eccrine spiradenoma
8407Sclerosing sweat duct carcinoma
8408Eccrine papillary adenocarcinoma
8409Eccrine poroma, malignant
8410Sebaceous adenocarcinoma
8413Eccrine adenocarcinoma
8420Ceruminous adenocarcinoma
8430Mucoepidermoid carcinoma
8440Cystadenocarcinoma, NOS
8441Serous cystadenocarcinoma, NOS
8450Papillary cystadenocarcinoma, NOS
8452Solid pseudopapillary carcinoma
8453Intraductal papillary-mucinous carcinoma, invasive
8460Papillary serous cystadenocarcinoma
8461Serous surface papillary carcinoma
8470Mucinous cystadenocarcinoma, NOS
8471Papillary mucinous cystadenocarcinoma
8480Mucinous adenocarcinoma
8481Mucin-producing adenocarcinoma
8482Mucinous adenocarcinoma, endocervical type
8490Signet ring cell carcinoma

8500Infiltrating duct carcinoma, NOS
8501Comedocarcinoma, NOS
8502Secretory carcinoma of breast
8503Intraductal papillary adenocarcinoma with invasion
8504Intracystic carcinoma, NOS
8508Cystic hypersecretory carcinoma
8510Medullary carcinoma, NOS
8512Medullary carcinoma with lymphoid stroma
8513Atypical medullary carcinoma
8514Duct carcinoma, desmoplastic type
8520Lobular carcinoma, NOS
8521Infiltrating ductular carcinoma
8522Infiltrating duct and lobular carcinoma
8523Infiltrating duct mixed with other types of carcinoma
8524Infiltrating lobular mixed with other types of carcinoma
8525Polymorphous low grade adenocarcinoma
8530Inflammatory carcinoma
8540Paget disease, mammary
8541Paget disease and infiltrating duct carcinoma of breast
8542Paget disease, extramammary (except Paget disease of bone(
8543Paget disease and intraductal carcinoma of breast
8550Acinar cell carcinoma
8551Acinar cell cystadenocarcinoma
8560Adenosquamous carcinoma
8562Epithelial-myoepithelial carcinoma
8570Adenocarcinoma with squamous metaplasia
8571Adenocarcinoma with cartilaginous and osseous metaplasia
8572Adenocarcinoma with spindle cell metaplasia
8573Adenocarcinoma with apocrine metaplasia
8574Adenocarcinoma with neuroendocrine differentiation
8575Metaplastic carcinoma, NOS
8576Hepatoid adenocarcinoma
8580Thymoma, malignant, NOS
8581Thymoma, type A, malignant
8582Thymoma, type AB, malignant
8583Thymoma, type B1, malignant
8584Thymoma, type B2, malignant
8585Thymoma, type B3, malignant
8586Thymic carcinoma, NOS
8588Spindle epithelial tumor with thymus-like element
8589Carcinoma showing thymus-like element
8600Thecoma, malignant
8620Granulosa cell tumor, malignant
8630Androblastoma, malignant
8631Sertoli-Leydig cell tumor, poorly differentiated
8634Sertoli-Leydig cell, poorly diffn, with heterologous elements
8640Sertoli cell carcinoma
8650Leydig cell tumor, malignant
8670Steroid cell tumor, malignant
8680Paraganglioma, malignant

8693Extra-adrenal paraganglioma, malignant
8700Pheochromocytoma, malignant
8710Glomangiosarcoma
8711Glomus tumor, malignant
8720Malignant melanoma, NOS (except juvenile melanoma(
8721Nodular melanoma
8722Balloon cell melanoma
8723Malignant melanoma, regressing
8728Meningeal melanomatosis
8730Amelanotic melanoma
8740Malignant melanoma in junctional nevus
8741Malignant melanoma in precancerous melanosis
8742Lentigo maligna melanoma
8743Superficial spreading melanoma
8744Acral lentiginous melanoma, malignant
8745Desmoplastic melanoma, malignant
8746Mucosal lentiginous melanoma
8761Malignant melanoma in giant pigmented nevus
8770Mixed epithelioid and spindle cell melanoma
8771Epithelioid cell melanoma
8772Spindle cell melanoma, NOS
8773Spindle cell melanoma, type A
8774Spindle cell melanoma, type B
8780Blue nevus, malignant
8800Sarcoma, NOS
8801Spindle cell sarcoma
8802Giant cell sarcoma (except of bone(
8803Small cell sarcoma
8804Epithelioid sarcoma
8805Undifferentiated sarcoma
8806Desmoplastic small round cell tumor
8810Fibrosarcoma, NOS
8811Fibromyxosarcoma
8812Periosteal fibrosarcoma
8813Fascial fibrosarcoma
8814Infantile fibrosarcoma
8815Solitary fibrous tumor, malignant
8830Malignant fibrous histiocytoma
8832Dermatofibrosarcoma, NOS
8833Pigmented dermatofibrosarcoma protuberans
8840Myxosarcoma
8850Liposarcoma, NOS
8851Liposarcoma, well differentiated
8852Myxoid liposarcoma
8853Round cell liposarcoma
8854Pleomorphic liposarcoma
8855Mixed liposarcoma
8857Fibroblastic liposarcoma
8858Dedifferentiated liposarcoma
8890Leiomyosarcoma, NOS

8891Epithelioid leiomyosarcoma
8894Angiomyosarcoma
8895Myosarcoma
8896Myxoid leiomyosarcoma
8900Rhabdomyosarcoma, NOS
8901Pleomorphic rhabdomyosarcoma, adult type
8902Mixed type rhabdomyosarcoma
8910Embryonal rhabdomyosarcoma, NOS
8912Spindle cell rhabdomyosarcoma
8920Alveolar rhabdomyosarcoma
8921Rhabdomyosarcoma with ganglionic differentiation
8930Endometrial stromal sarcoma, NOS
8931Endometrial stromal sarcoma, low grade
8933Adenosarcoma
8934Carcinofibroma
8935Stromal sarcoma, NOS
8936Gastrointestinal stromal sarcoma
8940Mixed tumor, malignant, NOS
8941Carcinoma in pleomorphic adenoma
8950Mullerian mixed tumor
8951Mesodermal mixed tumor
8959Malignant cystic nephroma
8960Nephroblastoma, NOS
8963Malignant rhabdoid tumor
8964Clear cell sarcoma of kidney
8970Hepatoblastoma
8971Pancreatoblastoma
8972Pulmonary blastoma
8973Pleuropulmonary blastoma
8980Carcinosarcoma, NOS
8981Carcinosarcoma, embryonal
8982Malignant myoepithelioma
8990Mesenchymoma, malignant
8991Embryonal sarcoma
9000Brenner tumor, malignant
9014Serous adenocarcinofibroma
9015Mucinous adenocarcinofibroma
9020Phyllodes tumor, malignant
9040Synovial sarcoma, NOS
9041Synovial sarcoma, spindle cell
9042Synovial sarcoma, epithelioid cell
9043Synovial sarcoma, biphasic
9044Clear cell sarcoma, NOS (except of kidney(
9050Mesothelioma, malignant
9051Fibrous mesothelioma, malignant
9052Epithelioid mesothelioma, malignant
9053Mesothelioma, biphasic, malignant
9060Dysgerminoma
9061Seminoma, NOS
9062Seminoma, anaplastic

9063Spermatocytic seminoma
9064Germinoma
9065Germ cell tumor, nonseminomatous
9070Embryonal carcinoma, NOS
9071Yolk sac tumor
9072Polyembryoma
9080Teratoma, malignant, NOS
9081Teratocarcinoma
9082Malignant teratoma, undifferentiated
9083Malignant teratoma, intermediate
9084Teratoma with malignant transformation
9085Mixed germ cell tumor
9090Struma ovarii, malignant
9100Choriocarcinoma, NOS
9101Choriocarcinoma combined with other germ cell elements
9102Malignant teratoma, trophoblastic
9105Trophoblastic tumor, epithelioid
9110Mesonephroma, malignant
9120Hemangiosarcoma
9124Kupffer cell sarcoma
9130Hemangioendothelioma, malignant
9133Epithelioid hemangioendothelioma, malignant
9140Kaposi sarcoma
9150Hemangiopericytoma, malignant
9170Lymphangiosarcoma
9180Osteosarcoma, NOS
9181Chondroblastic osteosarcoma
9182Fibroblastic osteosarcoma
9183Telangiectatic osteosarcoma
9184Osteosarcoma in Paget disease of bone
9185Small cell osteosarcoma
9186Central osteosarcoma
9187Intraosseous well differentiated osteosarcoma
9192Parosteal osteosarcoma
9193Periosteal osteosarcoma
9194High grade surface osteosarcoma
9195Intracortical osteosarcoma
9220Chondrosarcoma, NOS
9221Juxtacortical chondrosarcoma
9230Chondroblastoma, malignant
9231Myxoid chondrosarcoma
9240Mesenchymal chondrosarcoma
9242Clear cell chondrosarcoma
9243Dedifferentiated chondrosarcoma
9250Giant cell tumor of bone, malignant
9251Malignant giant cell tumor of soft parts
9252Malignant tenosynovial giant cell tumor
9260Ewing sarcoma
9261Adamantinoma of long bones
9270Odontogenic tumor, malignant

9290Ameloblastic odontosarcoma
9310Ameloblastoma, malignant
9330Ameloblastic fibrosarcoma
9342Odontogenic carcinosarcoma
9362Pineoblastoma
9364Peripheral neuroectodermal tumor
9365Askin tumor
9370Chordoma, NOS
9371Chondroid chordoma
9372Dedifferentiated chordoma
9380Glioma, malignant
9381Gliomatosis cerebri
9382Mixed glioma
9390Choroid plexus carcinoma
9391Ependymoma, NOS
9392Ependymoma, anaplastic
9393Papillary ependymoma
9400Astrocytoma, NOS
9401Astrocytoma, anaplastic
9410Protoplasmic astrocytoma
9411Gemistocytic astrocytoma
9420Fibrillary astrocytoma
9423Polar spongioblastoma
9424Pleomorphic xanthoastrocytoma
9430Astroblastoma
9440Glioblastoma, NOS
9441Giant cell glioblastoma
9442Gliosarcoma
9450Oligodendroglioma, NOS
9451Oligodendroglioma, anaplastic
9460Oligodendroblastoma
9470Medulloblastoma, NOS
9471Desmoplastic nodular medulloblastoma
9472Medullomyoblastoma
9473Primitive neuroectodermal tumor, NOS
9474Large cell medulloblastoma
9480Cerebellar sarcoma, NOS
9490Ganglioneuroblastoma
9500Neuroblastoma, NOS
9501Medulloepithelioma, NOS
9502Teratoid medulloepithelioma
9503Neuroepithelioma, NOS
9504Spongioneuroblastoma
9505Ganglioglioma, anaplastic
9508Atypical teratoid/rhabdoid tumor
9510Retinoblastoma, NOS
9511Retinoblastoma, differentiated
9512Retinoblastoma, undifferentiated
9513Retinoblastoma, diffuse
9520Olfactory neurogenic tumor

9521Olfactory neurocytoma
9522Olfactory neuroblastoma
9523Olfactory neuroepithelioma
9530Meningioma, malignant
9538Papillary meningioma
9539Meningeal sarcomatosis
9540Malignant peripheral nerve sheath tumor
9560Neurilemoma, malignant
9561Malig. peripheral nerve sheath tumor, rhabdomyoblastic difftn.
9571Perineurioma, malignant
9580Granular cell tumor, malignant
9581Alveolar soft part sarcoma
9590Malignant lymphoma, NOS
9591Malignant lymphoma, non-Hodgkin, NOS
9596Composite Hodgkin and non-Hodgkin lymphoma
9650Hodgkin lymphoma, NOS
9651Hodgkin lymphoma, lymphocyte-rich
9652Hodgkin lymphoma, mixed cellularity, NOS
9653Hodgkin lymphoma, lymphocyte depletion, NOS
9654Hodgkin lymphoma, lymphocyte depletion, diffuse fibrosis
9655Hodgkin lymphoma, lymphocyte depletion, reticular
9659Hodgkin lymphoma, nodular lymphocyte predominance
9661Hodgkin granuloma
9662Hodgkin sarcoma
9663Hodgkin lymphoma, nodular sclerosis, NOS
9664Hodgkin lymphoma, nodular sclerosis, cellular phase
9665Hodgkin lymphoma, nodular sclerosis, grade 1
9667Hodgkin lymphoma, nodular sclerosis, grade 2
9670Malignant lymphoma, small B lymphocytic, NOS
9671Malignant lymphoma, lymphoplasmacytic
9673Mantle cell lymphoma
9675Malignant lymphoma, mixed small and large cell, diffuse
9678Primary effusion lymphoma
9679Mediastinal large B-cell lymphoma
9680Malignant lymphoma, large B-cell, diffuse, NOS
9684Malignant lymphoma, large B-cell, diffuse, immunoblastic, NOS
9687Burkitt lymphoma, NOS
9689Splenic marginal zone B-cell lymphoma
9690Follicular lymphoma, NOS
9691Follicular lymphoma, grade 2
9695Follicular lymphoma, grade 1
9698Follicular lymphoma, grade 3
9699Marginal zone B-cell lymphoma, NOS
9700Mycosis fungoides
9701Sezary syndrome
9702Mature T-cell lymphoma, NOS
9705Angioimmunoblastic T-cell lymphoma
9708Subcutaneous panniculitis-like T-cell lymphoma
9709Cutaneous T-cell lymphoma, NOS
9714Anaplastic large cell lymphoma, T cell and Null cell type

9716Hepatosplenic (gamma-delta) cell lymphoma
9717Intestinal T-cell lymphoma
9718Primary cutaneous CD30+ T-cell lymphoproliferative disorder
9719NK/T-cell lymphoma, nasal and nasal-type
9727Precursor cell lymphoblastic lymphoma, NOS
9728Precursor B-cell lymphoblastic lymphoma
9729Precursor T-cell lymphoblastic lymphoma
9731Plasmacytoma, NOS
9732Multiple myeloma
9733Plasma cell leukemia
9734Plasmacytoma, extramedullary (not occurring in bone(
9740Mast cell sarcoma
9741Malignant mastocytosis
9742Mast cell leukemia
9750Malignant histiocytosis
9754Langerhans cell histiocytosis, disseminated
9755Histiocytic sarcoma
9756Langerhans cell sarcoma
9757Interdigitating dendritic cell sarcoma
9758Follicular dendritic cell sarcoma
9760Immunoproliferative disease, NOS
9761Waldenstrom macroglobulinemia
9762Heavy chain disease, NOS
9764Immunoproliferative small intestinal disease
9800Leukemia, NOS
9801Acute leukemia, NOS
9805Acute biphenotypic leukemia
9820Lymphoid leukemia, NOS
9823B-cell chronic lymphocytic leukemia/small lymphocytic lymphoma
9826Burkitt cell leukemia
9827Adult T-cell leukemia/lymphoma (HTLV-1 positive(
9832Prolymphocytic leukemia, NOS
9833Prolymphocytic leukemia, B-cell type
9834Prolymphocytic leukemia, T-cell type
9835Precursor cell lymphoblastic leukemia, NOS
9836Precursor B-cell lymphoblastic leukemia
9837Precursor T-cell lymphoblastic leukemia
9840Acute myeloid leukemia, M6 type
9860Myeloid leukemia, NOS
9861Acute myeloid leukemia, NOS
9863Chronic myeloid leukemia, NOS
9866Acute promyelocytic leukemia, t(15;17)(q22;q11-12(
9867Acute myelomonocytic leukemia
9870Acute basophilic leukemia
9871Acute myeloid leukemia with abnormal marrow eosinophils
9872Acute myeloid leukemia, minimal differentiation
9873Acute myeloid leukemia without maturation
9874Acute myeloid leukemia with maturation
9875Chronic myelogenous leukemia, BCR/ABL positive
9876Atypical chronic myeloid leukemia, BCR/ABL negative

9891Acute monocytic leukemia
9895Acute myeloid leukemia with multilineage dysplasia
9896Acute myeloid leukemia, t(8;21)(q22;q22(
9897Acute myeloid leukemia, 11q23 abnormalities
9910Acute megakaryoblastic leukemia
9920Therapy-related acute myeloid leukemia, NOS
9930Myeloid sarcoma
9931Acute panmyelosis with myelofibrosis
9940Hairy cell leukemia
9945Chronic myelomonocytic leukemia, NOS
9946Juvenile myelomonocytic leukemia
9948Aggressive NK-cell leukemia
9950Polycythemia vera
9960Chronic myeloproliferative disease, NOS
9961Myelosclerosis with myeloid metaplasia
9962Essential thrombocythemia
9963Chronic neutrophilic leukemia
9964Hypereosinophilic syndrome
9980Refractory anemia
9982Refractory anemia with sideroblasts
9983Refractory anemia with excess blasts
9984Refractory anemia with excess blasts in transformation
9985Refractory cytopenia with multilineage dysplasia
9986Myelodysplastic syndrome with 5q- syndrome
9987Therapy-related myelodysplastic syndrome, NOS
9989Myelodysplastic syndrome, NOS
9990well deffreniated adenocarcinoma

>---------- 6#Behaviour
0Benign
1Uncertain ben/malig
2In situ
3Malignant

>---------- 7#Sex
1Male
2Female
9Unknown

>---------- 8#Basis of Diagnosis
0Death Certificate Only
1Clinical only
2Clin. Invest./Ult Sound
3Surgery/Autopsy
4Laboratory test
5Cytology
6Histology of metastases
7Histology of primary
8Autopsy/Histology
9Unknown

>---------- 9#Tribe
00Foreign
01Mahas ãÍÓ
02Bederiya ÈÏíÑíÉ
03Burtie
04Bargow ÈÑÞæ
05Baruia
06Ahamda ÃÍÇãÏÉ
07shunabla
08Beja
09Junobia
0a Ben maen
0b Asilat
0c Balanda
10Barnow ÈÑäæ
11Bazia ÈÇÒíÇ
12Basharien ÈÔÇÑííä
13Butaheen ÈØÇÍíä
14Bagara ÈÞÇÑÉ
15Bein Jurar Èäì ÌÑÇÑ
16Bein Khalil Èäì Îáíá
17Bein Amer Èäì ÚÇãÑ
18Bein Hulbia Èäì åáÈíÉ
19Tama ÊÇãÉ
1a Agbat
1b Ashraf
1c Ethiobi
1d Somalia
20Torguim
21Taaisha ÊÚÇíÔÉ
22Tagalie
23Tungir
24Jaafarya ÌÚÇÝÑÉ
25Gaalya ÌÚáííä
26Gummoya ÌãæÚíÉ
27Gemaib ÌãíÚÇÈ
28Gwamaa ÌæÇãÚÉ
29Hassanya ÍÓÇäíÉ
30Hafurie
31Hlawien ÍáÇæííä
32Halfawe ÍáÝÇæííä
33Humer ÍãÑ
34Hawazma ÍæÇÒãÉ
35Dago ÏÇÌæ
36Dar Hamid ÏÇÑÍÇãÏ
37Dabasien ÏÇÈÇÓíä
38Dongla ÏäÇÞáÉ
39Dawgo
3a banawa

40Dwieh
41Denka ÏíäßÇ
42Rubatab ÑÈÇØÇÈ
43Rezegiat ÑÒíÞÇÊ
44Rashdy/a ÑÔÇíÏÉ
45Rufa'a ÑÝÇÚíÉ
46Rukabuia ÑßÇÈíÉ
47Zandie ÒÇäÌì
48Zgawa ÒÛÇæÉ
49Zubadya ÒÈíÏíÉ
50Slamat
51Salumie
52Shaigya ÔÇíÞíÉ
53Shukruia ÔßÑíÉ
54Shuluk Ôáß
55Shwehat
56Slehaib ÕáíÍÇÈ
57Ababdia ÚÈÇÈÏíÉ
58Abdelab ÚÈÏáÇÈ
59Arakia ÚÑßíÉ
5p Mowled ãæáÏ
60Agalie
61Awamria ÚæÇãÑíÉ
62Gubish
63Fadnya
64Fadlaib
65Flatia ÝáÇÊÉ
66For ÝæÑ
67Foung ÝæäÌ
68Guran ÞÑÚÇä
69Guzah
70Kahlya ßÇååíÉ
71Kbabish ßÈÇÈíÔ
72Kenana ßäÇäÉ
73Kenoz ßäæÒ
74Lahwuin áÍæííä
75Mabial
76Maganin ãÌÇäíä
77Maalya
78Mouhmduia
79Msauia
80Msaluit
81Musullamia ãÓáãíÉ
82Meserya ãÓíÑíÉ
83Mshaukia
84Maluia
85Mgarbia ãÛÑÈíÉ
86Mudoub
87Mnasier ãäÇÕíÑ
88Mwallad

89Maidoub
90Meema
91Nuba äæÈÉ
92Nuwier äæíÑ
93Hubbanya åÈÇäíÉ
94Hdandwia åÏäÏæÉ
95Hwawer åæÇæíÑ
96Hwsa åæÓÇ
97Gmraouy
98Latouka
99Unkown ÛíÑ ãÐßæÑ

>---------- 10#Region/District
10Khartoum ÇáÎÑØæã
11Algazira ÇáÌÒíÑÉ
12White Nile Çáäíá ÇáÃÈíÖ
13Blue Nile Çáäíá ÇáÇÒÑÞ
14Sinnar ÓäÇÑ
15River Nile äåÑ Çáäíá
16Northern ÇáÔãÇáíÉ
17Red Sea ÇáÈÍÑ ÇáÇÍãÑ
18Kasala ßÓáÇ
19Gadarif ÇáÞÖÇÑÝ
20N.Kordofan ÔãÇá ßÑÏÝÇä
21S.Kordofan ÌäæÈ ßÑÏÝÇä
22W.Kordofan ÛÑÈ ßÑÏÝÇä
23N.Darfor ÔãÇá ÏÇÑÝæÑ
24W.Darfor ÛÑÈ ÏÇÑÝæÑ
25S.Darfor ÌäæÈ ÏÇÑÝæÑ
26Bahr Algabal ÈÍÑ ÇáÌÈá
27Eastern Equatoria ÔÑÞ ÇáÇÓÊæÇÆíÉ
28W. Equatoria ÛÑÈ ÇáÇÓÊæÇÆíÉ
29Unity ÇáæÍÏÉ
30Warab æÇÑÇÈ
31Upper Nile ÃÚÇáì Çáäíá
32Gongoli ÌæäÞáì
33Bohairat ÇáÈÍíÑÇÊ
34N.Bahr Algazal ÔãÇá ÈÍÑ ÇáÛÒÇá
35W.Bahr Algazal ÛÑÈ ÈÍÑ ÇáÛÒÇá
36Outside Sudan
1036Khartoum ÇáÎÑØæã
1037Gabal Awlia ÌÈá ÇæáíÇÁ
1038Bahri ÈÍÑì
1039Sharg Alnil ÔÑÞ Çáäíá
1040Umdurman Çã ÏÑãÇä
1041Karari ßÑÑì
1042Umbada Çã ÈÏÉ
1143Great madani ãÏäì ÇáßÈÑì
1144South Gezira ÌäæÈ ÇáÌÒíÑÉ
1145Almanagil ÇáãäÇÞá

1146Alkamlin ÇáßÇãáíä
1147Alhasahesa ÇáÍÕÇÍíÕÇ
1148East Gezira ÔÑÞ ÇáÌÒíÑÉ
1149Um algura Çã ÇáÞÑì
1250Kosti ßæÓÊì
1251Rabak ÑÈß
1252Aldwam ÇáÏæíã
1253Algetana ÇáÞØíäÉ
1354Aldamazin ÇáÏãÇÒíä
1355Alrusars ÇáÑÕíÑÕ
1356Baw ÈÇæ
1357Gasan ÞíÓÇä
1358Alkurmuk ÇáßÑãß
1459Senga ÓäÌÉ
1460Senar ÓäÇÑ
1461Aldindir ÇáÏäÏÑ
1562Atbra ÚØÈÑÉ
1563Aldamar ÇáÏÇãÑ
1564Shandi ÔäÏì
1565Almatama ÇáãÊãÉ
1566Barbar ÈÑÈÑ
1567Abu Hamad ÇÈæÍãÏ
1668Wadi Halfa æÇÏì ÍáÝÇ
1669Dongola ÏäÞáÇ
1670Aldaba ÇáÏÈÉ
1671Marawi ãÑæì
1772Port Sudan ÈæÑÊÓæÏÇä
1773Sinkat ÓäßÇÊ
1774Toker ØæßÑ
1775Halayb ÍáÇíÈ
1876Kassala ßÓáÇ
1877Algash ÇáÞÇÔ
1878Hamashkorab åãÔßæÑíÈ
1879Nahr atbra ääÑ ÚØÈÑÉ
1880Setate ÓÊíÊ
1981Algadarif ÇáÞÖÇÑÝ
1982Algalabat ÇáÞáÇÈÇÊ
1983Alfashga ÇáÝÔÞÉ
1984Alfaw ÇáÝÇæ
1985Alrahad ÇáÑåÏ
2086Shakan ÔíßÇä
2087Umrawaba Çã ÑæÇÈÉ
2088Bara ÈÇÑÇ
2089Sodari ÓæÏÑì
2090Gabrat alshash ÌÈÑÉ ÇáÔíÎ
2191Kadogli ßÇÏæÞáì
2192Aldalang ÇáÏääÌ
2193Alrashad ÇáÑÔÇÏ
2194Abugebaha ÇÈæÌÈíåÉ
2195Talodi ÊÇáæÏì

2201Lagawa áÞÇæÉ
2296Abyai ÇÈíí
2297Alnohood Çáäåæï
2298Gebash ÛÈíÔ
2299Alsalam ÇáÓáÇã
2302Alfashir ÇáÝÇÔÑ
2303Maleet ãáíØ
2304Kutom ßÊã
2305Alteena ÇáØíäÉ
2306Kabkabia ßÈßÇÈíÉ
2307Umkadada Çã ßÏÇÏÉ
2408Algenana ÇáÌäíäÉ
2409Zalengi ÒÇääÌì
2410Wad salih æÏ ÕÇáÍ
2411Gabal Mara ÌÈá ãÑÉ
2512Niala äíÇáÇ
2513Kas ßÇÓ
2514Buram ÈÑÇã
2515Tulus ÊáÓ
2516Aldeian ÇáÏ íÇä
2517Ad Alfursan ÚÏ ÇáÝÑÓÇä
2518Rehad alburdi ÑåíÏ ÇáÈÑÏì
2519Adeela ÚÏíáÉ
2620Bahr Algabal ÈÍÑ ÇáÌÈá
2621Kabokabi ßÇÈæßÇÈì
2622Nahr Yai äåÑ íÇì
2623Trkaka ÊÑßÇßÇ
2724Kaboita ßÈæíÊÇ
2725Shakdoom ÔßÏæã
2726Amatong ÃãÇÊæäÌ
2727Magwi ãÞæì
2828Mondri ãäÏÑì
2829Yambu íÇãÈíæ
2830Mareedi ãÑíÏì
2831Tombra ØãÈÑÉ
2932Rebcona ÑÈßæäÇ
2933Myom ãíæã
2934Allar ÇááÑ
2935Fareeng ÝÇÑäÌ
3036Gogrial ÞæÞÑíÇá
3037Altong ÇáÊæäÌ
3038Warab æÇÑÇÈ
3039Nahr Allol äåÑ Çááæá
3140Tonba ÊæäÈÉ
3141Alsobat ÇáÓæÈÇØ
3142Arnk ÇÑäß
3143Maloot ãÇáæÊ
3144Almaban ÇáãÇÈÇä
3145Mayot ãÇíæÊ
3146Balat ÈÇáÇÊ

3147Fashoda ÝÔæÏÉ
3248Fam Alzaraf Ýã ÇáÒÑÇÝ
3249Akobo ÇßæÈæ
3250Nahr Ateem äåÑ ÃÊíã
3251Bor ÈæÑ
3252Wat æÇÊ
3253Nahr Albebor äåÑ ÇáÈíÈæÑ
3254Ayod ÇíæÏ
3355Nahr yarol äåÑ íÑæá
3356Alaalyab ÇáÚÇáíÇÈ
3357Rumbak ÑãÈíß
3358Shobat ÔæÈíÊ
3459Aryat ÇÑíÇÊ
3460Malook ãÇáæß
3461Aweel Çæíá
3462Wangook æÇäÌæß
3563Nahr Algoor äåÑ ÇáÌæÑ
3564Wao æÇæ
3565Raga ÑÇÌÇ
3666Egypt
3667Ethiopia
3668Zaeir
3669Emarat
3670Somali

>---------- 11#Occupation
10Accountant ãÍÇÓÈ
11Baker ÝÑÇä - ÎÈÇÒ
12Barber ÍáÇÞ
13Blacksmith ÍÏÇÏ
14Lawyer ãÍÇãì
15Butcher ÌÒÇÑ
16Carpenter äÌÇÑ
17Carver äÞÇÔ
18Cashier Çãíä ÕäÏæÞ - ßÇÔíÑ
19Cook ØÈÇÎ
20Driver ÓÇÆÞ
21Electrician Ýäì ßåÑÈÇÁ
22Engineering ãåäÏÓ
23Firefighter ÑÌá ÇØÝÇÁ - ÇØÝÇÆì
24Garbage Collector ÌÇãÚ ÇáäÝÇíÇÊ
25Farmer ãÒÇÑÚ
26Guard ÍÇÑÓ
27Goldsmith ÕÇÆÛ
28Herdsman ÑÇÚì
29House wife ÑÈÉ ãäÒá
30Illustrator ÑÓÇã
31Journalist ÕÍÝì
32Lecturer ãÍÇÖÑ ÌÇãÚì
33Librarian Çãíä ãßÊÈÉ

34Machinist ãíßÇäíßì
35Medical Doctor ØÈíÈ
36Merchant ÊÇÌÑ
37Midwife ÞÇÈáÉ
38Musician ÚÇÒÝ - ãæÓíÞì
39Nurse ããÑÖ / ããÑÖÉ
40Pharmacist ÕíÏáì
41Photographer ãÕæÑ
42Plumber ÓÈÇß
43Police ÈæáíÓ
44Receptionist ÚÇãá ÇáÇÓÊÞÈÇá
45Sailor ÈÍÇÑ / ãáÇÍ
46Soldier ÚÓßÑì / ÌäÏì
47Secretary ÓßÑÊíÑ / ÓßÑÊíÑÉ
48Student ØÇáÈ / ØÇáÈÉ
49Teacher ãÚáã / ãÚáãÉ
50Veterinarian ÈíØÑì
51Free worker ÇÚãÇá ÍÑÉ
52Un employee ÚÇØá /áÇ íÚãá
53Not mentioned ÛíÑ ãÚÑæÝ
54gov employee
55emam
56labour ÚÇãá
57Child
58Retired emploee

>---------- 12#Hospital
100Khartoum Techniq Hospital
101Omdurman Techniq Hospital
102Bahry Techniq Hospital
103Soba Techniq Hospital
104Al-Shaab Techniq Hospital
105Khartoum ENT Hospital
106Omdurman ENT Hospital
107Bahry ENT Hospital
108Saad Abualala hospital
.... 109
110Khartoum Dental Hospital
111IBN Siha Hospital
112Abu-Anga Techniq Hospital
113Eye Specialization Hospital Khartoum
114Eye Specialization Hospital Bahry
115Eye Specialization Hospital Omdurman
116Omdurman Tropical Hospital
117Chinese Teachniq Hospital.Omdurman
118Omdurman Militry Hospital
119Ibrahim Malik Hospital.Khartoum
120Bashier Teachniq Hospital.Khartoum
121Turkish Hospital Khartoum
122Academy Hospital Khartoum

123Al-Walidain Hospital Omdurman
124Al-Soudi Hospital Omdurman
125Omdurman Paediatric Hospital
126Al-Boluk Hospital Omdurman
127Al-Naw Hospital Omdurman
128Ombadda Hospital Omdurman
129Other Hospital
130INMO
131RICK
132Uuknown
133Private clinic
134Private hospital
135Police hospital
136Omdurman dermatology hospital
137Bahri dermatology hospital
138Gafar Ibn Ouf Paediatric Hospital
139Omdurman Maternaty hospital
140Alribat Hospital
141Alengaz hospital
142Emarat
143Cairo
144Khartoum dermatology hospital
145Jordan
146Saudi arabia
147Lebia
148Somalia

>---------- 13#Department
10Surgery
11Orthopaedics
12Urology
13Dental Surg.
14Opthalmology
15ENT
20Medecine
21Paediatrics
22Dermatology
30Obs / Gynae
31Gynaecology
80Other
99Unknown

>---------- 14#Status
0Dead
1Alive

>---------- 15#Yes/No
0No
1Yes

>---------- 16#Stage Cervix
00TX
0 01
02I
03IA
04IA1
05IA2
06IB
07IB1
08IB2
09II
10IIA
11IIB
12III
13IIIA
14IIIB
15IVA
16IVB

>---------- 17#Gleason Grade
1 1
2 2
3 3
4 4
5 5

>---------- 18#SEER
0in situ
1loc. only
2reg. ext. only
3reg. Lymph. only
4reg. ext.+lymph.
5Reg. NOS
7Distant
9Unk. ext/met.

>---------- 19#Marital Status
0 اعزب
1 متزوج
2 ارمل
3 مطلق
9 غير محدد