



جامعة السودان للعلوم والتكنولوجيا

كلية علوم الحاسوب وتقانة المعلومات

قسم علوم الحاسوب

**تصنيف بيانات المصفوفات الدقيقة باستخدام**

**الشبكات العصبية ذات التعليم المتعمق**

دراسة حالة: بيانات سرطان الثدي

أكتوبر 2016

مشروع مقدم كأحد متطلبات الحصول على بكالوريوس الشرف في كلية علوم

الحاسوب



بسم الله الرحمن الرحيم

جامعة السودان للعلوم والتكنولوجيا

كلية علوم الحاسوب وتقانة المعلومات

تصنيف بيانات المصفوفات الدقيقة باستخدام

الشبكات العصبية ذات التعليم المتعمق

دراسة حالة: بيانات سرطان الثدي

أكتوبر 2016

إعداد :

رنين محمد أحمد ميرغني

مرح محمد أحمد

هند يوسف أحمد

التوقيع :

.....

إشراف :

د. هويدا علي عبدالقادر

/أكتوبر/ 2016







# الحمد لله

الحمد لله ...

الحمد لله الملك الجليل، نحمده كلما نُطق بحمده وقيل، المنزّه عن النظير والعديل  
المنعم بقبول القليل، المتكرم بإعطاء الجزيل، تقدّس عمّا يقول أهل التعطيل، وتعالى عمّا  
يعتقد أهل التمثيل، نصبه للعقل على وجوده أوضح دليل، وهدى إلى وجوده بَيّن سبيل،  
وجعل للحسن خطأً إلى مثله يميل .

نحمد الله الذي فضلنا على كثيرٍ من خلقه وانعم علينا بنعمة العقل وزودنا بأدوات  
العلم ومهد لنا الطريق للإتمام هذا البحث العلمي المتواضع الذي نأمل من المولى عز  
وجل ان يكون ذا جدوى لما أخرج له .

# إهداء

بسم الله الرحمن الرحيم

قُلْ أَعْمَلُ لِلْوَلِيِّ سَيْرَى اللّٰهُمَّ لَكُمْ وَرَسُولِ الْوَالِدِ وَمِنْ دُونَ

صدق الله العظيم

إلهي لا يطيب الليل إلا بشرك ولا يطيب النهار إلا بطاعتك ولا تطيب اللحظات إلا بذكرك  
ولا تطيب الآخرة إلا بعفوك ولا تطيب الجنة إلا برويتك .

الله جل جلاله

الى من بلغ الرسالة وأدى الأمانة ونصح الأمة، إلى نبي الرحمة ونور العالمين.

سيدنا محمد صلى الله عليه وسلم

الى من كللهم الله بالهبة والوقار، إلى من علمونا العطاء بدون انتظار، إلى من نحمل  
أسماءهم بكل افتخار نرجو من الله أن يمد في أعماركم لتروثماراً قد حان قطافها بعد  
طول انتظار وستبقى كلماتكم جوماً نهتدي بها اليوم وفي الغد وإلى الأبد .

أباؤنا الأعزاء

إلى ملائكتنا في الحياة، إلى معنى الحب وإلى معنى الحنان والتفاني، إلى بسمة الحياة،  
وسر الوجود إلى من كان دعائهن سر نجاحنا وحنانهن بلسم جراحنا إلى أغلى الحبايب.

أمهاتنا الحبيبات



# شكر و عرفان

ومن حق النعمة الذكر، وأقل جزاء للمعروف الشكر .

بعد شكر المولى عز وجل المتفضل بجليل النعم وعظيم الجزاء، يجدر بنا أن نتقدم  
ببالغ الإمتنان وجزيل العرفان إلى كل من وجهنا وعلمنا وأخذ بيدنا في سبيل إنجاز هذا  
البحث.

كل الشكر و التقدير و الإحترام لمشرفتنا الأستاذة الدكتورة : هويدا علي عبدالقادر  
التي لم تبخل بتقديم يد العون لنا والتي قومت وتابعت وصوبت بحسن إرشادها لنا في كل  
مراحل البحث، والتي وجدنا في توجيهاتها حرص المعلم التي تؤتي ثمارها الطيبة بإذن  
الله .

كما نحمل الشكر والعرفان إلى كل من أمدنا بالعلم والمعرفة وأسدى لنا النصح  
والتوجيه ونتوجه بذلك الشكر إلى الاء بشير، إيمان محمد .

الشكر موصول إلى كل من ساندنا بدعواته الصادقة أو تمنياته الخالصة نشكرهم  
جميعا و نتمنى من الله عز وجل أن يجعل ذلك في موازين حسناتهم .

وكل من ساعدنا ولم نذكره فله منا العتبي حتى يرضى، وليعلم أن ذلك إنما هو سهو  
ونسيان من شخوصنا الضعيفة وليس تقليلاً من قدرته بأي حال من الأحوال .

# المستخلص

يهدف هذا البحث إلى استخدام الشبكات العصبية ذات التعليم المتعمق في تصنيف البيانات الحيوية خاصة المصفوفات الدقيقة وذلك للإستفادة من خاصية التعليم المتعمق في تقليل مدخلات وخصائص الشبكة العصبية.

يتم تصنيف بيانات المصفوفات الدقيقة والبحث عن الجينات الأكثر تأثيراً في تشخيص سرطان الثدي وذلك بإستخدام Feed forward Back propagation Neural Networks في تطبيق الماتلاب لعملية تقليل عدد الجينات من أجل إستخدامها في بعض المعلومات الحيوية لمراقبة التغيرات التي تحدث في كثير من الأمراض وخاصة السرطان وتوفير طرق علاج فعالة .

# ABSTRACT

This research aims to use the Deep learning Neural Network in addition to classify Bioinformatics specially microarray data. Therefore to taking the advantages of the deep learning to reduce inputs and features of the neural network.

We classified microarray data and identified genes that mostly affects in breast cancer diagnosis using Feed forward Back propagation Neural Networks in MATLAB for eliminating the number of features to help better observation for the mutations that could happen especially in cancer, and providing a better treatment methods.

## فهرس المصطلحات

Feed forward Back propagation Neural Networks	شبكة الإنتشار الخلفي أمامية التغذية
Bioinformatics	المعلومات الحيوية
NCBI	المركز الوطني لمعلومات التكنولوجيا الحيوية
Gene	الجين
RNA	الحمض الرايبوزي
DNA	الحمض النووي
mRNA	الحمض الرايبوزي الرسول
rRNA	التعلم المراقب
tRNA	الحمض الرايبوزي الناقل
<u>Protein</u>	البروتين
Nucleotide	نيوكليوتيدات
Microarray	المصفوفات الدقيقة
Features	مستقبلات
Protein microarray	مصفوفة البروتين الدقيقة
Antibody	أجسام مضادة
Biomarker	مؤشر حيوي
DNA microarray	الحيوية DNA مصفوفة
Tissue microarray	مصفوفة النسيج الحيوية
Machine Learning	تعلم الآلة
Supervised Learning	التعلم المراقب
Classification	التصنيف

Regression	الإحدار
Unsupervised Learning	التعلم الغير مراقب
Clustering	التجميع
Reinforcement learning	التعلم التعزيزي
Pattern Recognition	التعرف على الانماط
OCR	التعرف الضوئي على الحروف
Neural Network	الشبكة العصبية الاصطناعية
Deep Learning	التعلم المتعمق

## فهرس الأشكال

رقم الشكل	العنوان	رقم الصفحة
1.2	الخلية	8
2.2	شكل ومكونات الأحماض	10
3.2	بنية البروتين	12
4.2	رسم تمثيلي لمختلف أجزاء الجين	14
5.2	طريقة تحليل الحمض النووي	15
6.2	مصفوفة إختبار	18
7.2	التعلم المراقب والغير مراقب	24
8.2	الشبكة العصبية الطبيعية	31
9.2	الشبكة العصبية الاصطناعية	32
10.2	شبكة تعلم متعمق	36
1.4	موقع GEO	49
2.4	عدد العينات	50
3.4	Excel مجموعة البيانات في ملف	51
4.4	حالة العينة (b , c)	52
5.4	شجرة التدريب	53
6.4	شجرة الإختبار	54
7.4	مخرجات الشبكة	55

فهرس الجداول :

رقم الصفحة	العنوان	رقم الجدول
11	الفرق بين RNA و DNA	1.2

# فهرس المحتويات

رقم الصفحة	الموضوع
أ	الآية
ب	الحمد لله
ت	الإهداء
ث	الشكر والعرفان
ج	المستخلص
ح	Abstract
خ - د	شرح المصطلحات
ذ	فهرس الأشكال
ر	فهرس الجداول
ز - س - ش - ص - ض - ط	فهرس المحتويات
	<b>المقدمة</b>
	<b>الباب الأول</b>
2	1.1 المقدمة
2	2.1 مشكلة البحث
2	3.1 فرضية البحث
2	4.1 هدف البحث
3	5.1 أهمية البحث
3	6.1 حدود البحث
3	7.1 محتوى البحث
	<b>الباب الثاني</b>
7	1.2 المعلومات الحيوية



7	1.1.2 مقدمة
7	2.1.2 تعريف المعلومات الحيوية
8	2.1.3 المصطلحات المتعلقة بالمعلومات الحيوية

13	2.1.4 نبذة تاريخية عن المعلومات الحيوية
15	2.1.5 أهداف المعلومات الحيوية
16	6.1.2 تطبيقات المعلومات الحيوية
16	7.1.2 المشاكل التي تواجه المعلومات الحيوية
17	2.2 المصفوفات الدقيقة
17	1.2.2 مقدمة
17	2.2.2 تعريف المصفوفات الدقيقة
19	3.2.2 إستخدامات المصفوفات الدقيقة
20	4.2.2 أنظمة المصفوفات الدقيقة
21	5.2.2 مشكلة المصفوفة الدقيقة
23	3.2 تعلم الآلة
23	1.3.2 مقدمة
23	2.3.2 خوارزميات التعلم
25	3.3.2 أمثلة خوارزميات تعلم الآلة
25	4.3.2 تطبيقات تعلم الآلة
26	4.2 التعرف على الأنماط
26	1.4.2 مقدمة
26	2.4.2 مجالات التعرف على الأنماط
27	3.4.2 الخطوات العامة لنظام التعرف على الأنماط
28	2.4.4 طرق التعرف على الأنماط
29	5.4.2 تطبيقات التعرف على الأنماط
30	5.2 الشبكات العصبية
30	1.5.2 مقدمة
30	2.5.2 تعريف الشبكات العصبية
31	3.5.2 خصائص الشبكات العصبية
31	4.5.2 طريقة معالجة المعلومات

32	5.5.2 نماذج الشبكات العصبية
33	6.5.2 أنواع الشبكات العصبية
33	7.5.2 تطبيقات الشبكات العصبية
34	6.2 التعلم المتعمق
34	1.6.2 مقدمة
34	2.6.2 تعريف التعلم المتعمق
35	3.6.2 تقنيات التعلم العميق
37	4.6.2 صنوف التعلم المتعمق
37	5.6.2 خوارزميات التعلم المتعمق
39	<b>الدراسات السابقة</b>
	<b>الباب الثالث</b>
43	1.3 مقدمة
43	2.3 خصائص الماتلاب
43	3.3 استخدامات الماتلاب
44	3.4 نظام الماتلاب
44	1.4.3 لغة الماتلاب
44	2.4.3 بيئة عمل الماتلاب
44	3.4.3 معالجة الرسومات
44	4.4.3 الدوال الحسابية
44	5.3 واجهة المستخدم الرسومية
45	1.5.3 طريقة عمل واجهة المستخدم الرسومية
45	6.3 مميزات الماتلاب
46	7.3 عيوب الماتلاب
	<b>الباب الرابع</b>
	<b>التطبيق العملي</b>

49	1.4 المعالجة الأولية للبيانات
53	2.4 تحميل البيانات
	<b>الباب الخامس</b>
	<b>النتائج والتوصيات</b>
56	1.5 المقدمة
56	2.5 النتائج
57	3.5 التوصيات
57	4.5 الخاتمة

# المقدمة

# الباب الأول



# المقدمة

إن التطور الكبير الذي نشهده في مختلف مجالات العلوم والمعرفة والنتائج عن بحوث الكمبيوتر في الكثير من المجالات العلمية والصناعية والإقتصادية وغيرها فرض إهتماماً واضحاً بطبيعة وتركيبية الإنسان وذلك من خلال إيجاد خصائص معينة تساعد على إكتشاف الأمراض، ولأن الإنسان يسعى دائماً بكل جد وجهد لإكتشاف مزيد من الأسرار التي يجهلها في نشأ الحياة وتطور الكائنات الحية والبحث عن تفسير العديد من الظواهر الحيوية التي ظل عاجزاً عنها، كان لابد من إستخدام الكمبيوتر وتقنياته من أجل إنتاج المعادلات والخوارزميات الرياضية لتحليل وتنظيم المعلومات الحيوية الناتجة عن دراسات وأبحاث الآف المراكز البحثية المنتشرة حول العالم في مختلف الدول والعمل على ربط تلك البيانات وتنسيقها وتوفيرها للجميع لإستكمال الدراسات والبحوث عنها وإضافة المزيد إليها ومن هنا نشأ علم المعلومات الحيوية .

عرفت المعلومات الحيوية على أنها تطبيق التكنولوجيا الحاسوبية والمعلوماتية في إدارة المعلومات البيولوجية وبشكل محدد هي علم تطوير قواعد بيانات وخوارزميات حاسوبية لتسهيل وتسريع الأبحاث البيولوجية .

## 2.1 مشكلة البحث

تحتوي المصفوفات الدقيقة على كمية ضخمة من البيانات مما يصعب على الشخص العادي إجراء عمليات تصنيفها يدوياً .

## 3.1 فرضية البحث

الشبكات العصبية الإصطناعية المدربة على بيانات المصفوفات الدقيقة تتنبأ بعدد من المعلومات المفيدة التي تساعد في المستقبل على توفير فهم أفضل لكميات البيانات الحيوية الضخمة مما سوف يعين في المجال الطبي في حقول صنع الأدوية بصورة أفضل و تطوير طرق علاج فعالة وأيضاً في مراقبة التغيرات التي قد تحدث في كثير من الأمراض ومنها السرطان .

## 4.1 هدف البحث

يهدف هذا البحث إلى تصميم نموذج لتصنيف بيانات المصفوفات الدقيقة لمرضى سرطان الثدي للمساعدة في التشخيص آلياً وتقليل أو تقليل خصائص مدخلات الشبكة العصبية .



## 5.1 أهمية البحث

يساعد علماء الأحياء في الإستخدام الأمثل للتكنولوجيا في تصنيف البيانات الضخمة خاصة مرضى السرطان مما يساعد في العلاج.

## 6.1 حدود البحث

- صعوبة الوصول للبيانات.
- عدم وضوح البيانات.
- الحجم الكبير للبيانات.

## 7.1 محتوى البحث

**الباب الأول :** يحتوي على المقدمة، المشكلة، الفرضية، الهدف، الأهمية، حدود البحث والمحتوى .

**الباب الثاني :** الإطار النظري .

**الباب الثالث :** التقنية المستخدمة .

**الباب الرابع :** التطبيق العملي .

**الباب الخامس :** النتائج والتوصيات والملاحق .

# الباب الثاني

## الإطار النظري و الدراسات السابقة

# الفصل الأول

1.1.2 المعلومات الحيوية

2.1.2 المصفوفات الدقيقة

# المعلومات الحيوية

## 1.1.1.2 مقدمة

شهد البحث العلمي تطوراً ليس له مثيل في تاريخ البشرية حيث أن النظريات والمبادئ التي إستحدثت قبل سنتين قد أصبحت الآن في خبر كان وإستبدلت أو عدلت لتصبح أكثر فعالية، ومن نتائج هذا التطور هو إدخال تكنولوجيا الحاسوب و ذلك يعود لسببين : الأول هو مرور العالم من زمن كانت المشكلة فيه الحصول على المعلومات إلى زمن يحتوي على فيض من المعلومات بحيث أن أصغر مختبر ينتج في من البيانات مما يجعل التحليل اليدوي لهذه البيانات ضرباً من ضروب أو تيرابايت إسبوع واحد جيغابايت المستحيل ولهذا إستلزم إنشاء برامج وخوارزميات تقوم بتحليل هذه البيانات والمساعدة على فهمها، والسبب الثاني هو سهولة التخزين والبحث والنسخ حيث يمكن تقصي المعلومات السابقة في ظرف ثواني بالإضافة إلى سهولة تحديثها وعمل عدة نسخ منها، وعلم الأحياء كغيره من العلوم لا يشذ عن غيره في هذا المجال حيث مع البدايات الأولى لظهور الحاسوب أبدى العلماء تلهفهم لإستغلال هذه التكنولوجيا، كما أمنت الرياضيات للمعلومات الحيوية السير بمعدلات سريعة، وتحقيق قفزات تتلاءم مع حجم المعلومات المتضاعفة، ومن أجل أن تتكامل هذه المعلومات وتتصافر لتشكل صورة شاملة عن نشاطات الخلايا الحية مما ساعد الباحثين في الكشف عن أسباب تغير هذه النشاطات في حالات المرض المختلفة .

## 2.1.1.2 تعريف المعلومات الحيوية

المعلومات الحيوية هي تطبيق أحدث تقنيات الرياضيات التطبيقية والمعلوماتية والإحصاء وعلوم الحاسوب لمعالجة وتحليل البيانات البيولوجية ومشاكلها، وهو أحد التخصصات العلمية التي تتطور سريعاً [1] .

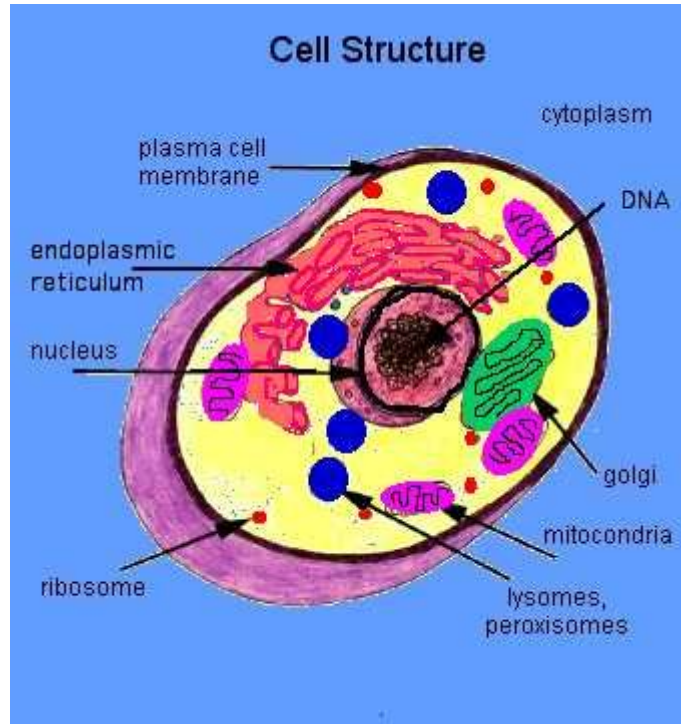
تنقسم المعلومات الحيوية الى ثلاثة أقسام رئيسية :

- تحليل و تفسير الأنماط المختلفة من البيانات التي تتضمن سلاسل الحموض الأمينية .
- تطوير خوارزميات جديدة وتقنيات إحصائية تساعد في تحصيل المعلومات من مجموعات ضخمة من البيانات .
- تطوير وتنفيذ أدوات تساعد على إدارة فعالة للأنماط المختلفة من المعلومات .

## 3.1.1.2 المصطلحات المتعلقة بالمعلومات الحيوية

### 1.3.1.1.2 الخلية

تتكون الكائنات الحية جميعها في أساس تركيبها من مجموعة لا نهائية من الخلايا المتجددة. تعرف الخلية بأنها الوحدة التركيبية الأساسية للإنسان والحيوان والنبات، كما تصنف بأنها المسؤولة عن التنوع العملي والوظيفي لمختلف أنواع الأحياء، بحيث تتنوع أشكالها وأحجامها وطريقة تركيبها مع بعضها البعض لتشكيل عضواً أو نسيجاً لتحقيق هدف محدد في الجسم. تتعرض الخلايا للتلف والضعف إلا أنها تعود للنمو وتكوين خلايا جديدة لتعويض التالف منها، إلا أنها قد تعجز عن ذلك في بعض الحالات مما يسبب ضموراً في أحد أعضاء الجسم وخلاً في تأديته لهدفه الوظيفي[2].



الشكل 1.2 : يوضح الخلية

## 1.1.3.1.1.2 تصنيفات الخلايا

تتعدد التقسيمات في تصنيف الخلايا منها ما يلي :

### 1.1.1.3.1.1.2 تصنيف الخلايا حسب عددها :

- وحيدة الخلية : يتميز هذا النوع بعدم قدرته على الإنقسام فيموت الكائن الحي وحيد الخلية في حال موت الخلية وتقوم هذه الخلية بجميع الوظائف، مثل خلية الأميبا.
- عديدة الخلايا : يتميز هذا النوع بقدرته على الإنقسام والتكاثر وتعويض الخلايا التالفة، كما تتطور لتكوين عدد من الأعضاء مثل الخلايا المكونة لجسم الإنسان.

### 2.1.1.3.1.1.2 تصنيف الخلايا حسب تطورها :

- خلية بدائية النواة.
- خلية حقيقية النواة.

### 3.1.1.3.1.1.2 تصنيف الخلايا حسب الكائنات الحية الموجودة فيها :

- الخلايا النباتية.
- الخلايا الحيوانية.

## 2.3.1.1.2 الجين

تحتوي الخلية على حوالي 25000-35000 جينات. الجين هو الذي يحمل المعلومات التي تحدد السمات أو الصفات التي يرثها الإنسان من الأباء يختلف حجم الجينات بحسب قواعد الحمض النووي التي يمكن أن تكون بالمئات أو قد يصل عددها إلى 2 مليون.

كل شخص لديه نسختين من كل جين مورثة من الأباء، معظم الجينات لا تختلف من شخص لآخر باستثناء عدد قليل من الجينات.

## 3.3.1.1.2 الأحماض النووية

يوجد نوعين من الأحماض النووية :

### DNA 1.3.3.1.1.2

الحمض النووي هو اختصار للحمض النووي الرايبوزي منقوص الأوكسجين

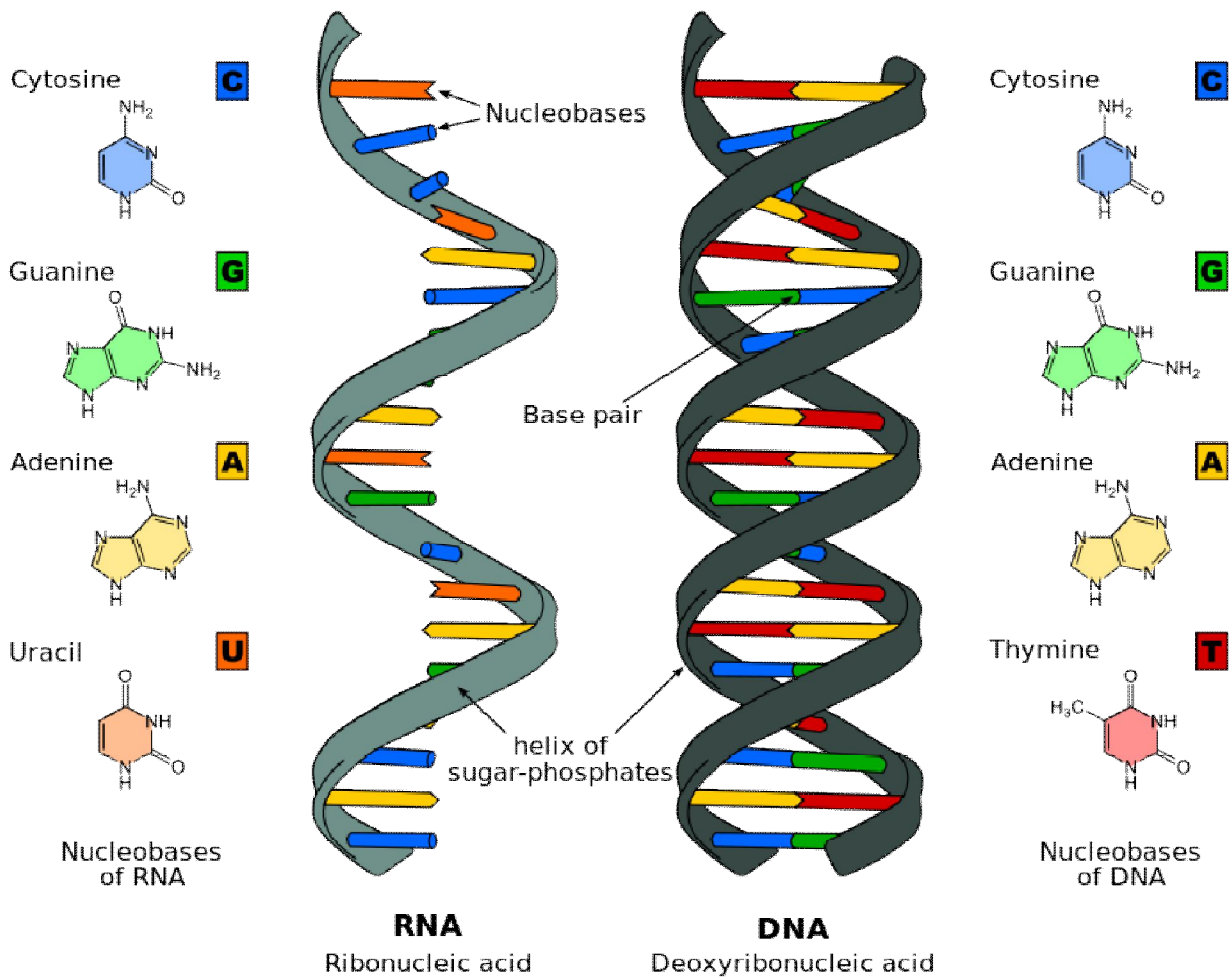


يحمل DNA الإختلاف بين البشر من حيث الجنس والشكل واللون وقد تمكن العالمان جيمس واتسون وفرانسيس كريك من إكتشاف الشكل الحلزوني ل DNA وتبين أنه مركب من سلاسل من الأحماض النووية، وأدى ذلك إلى التعرف على كيفية تخزين المعلومات الوراثية وحفظها وكيفية نقلها من جيل لآخر.

## RNA 2.3.3.1.1.2

ينقسم إلى ثلاثة أنواع :

- mRNA أو RNA الرسول : يقوم بنقل الشفرة الوراثية من الجينات في النواة إلى الرايبوسومات ليتم تصنيع البروتينات المختلفة.
- rRNA أو الرايبوسومي : يقوم بإنتاج الرايبوسومات داخل نواة الخلية.
- tRNA أو RNA الناقل : يقوم بنقل الأحماض الأمينية لإستخدامها في عملية بنائ البروتينات.



الشكل 2.2 : يوضح شكل ومكونات الأحماض

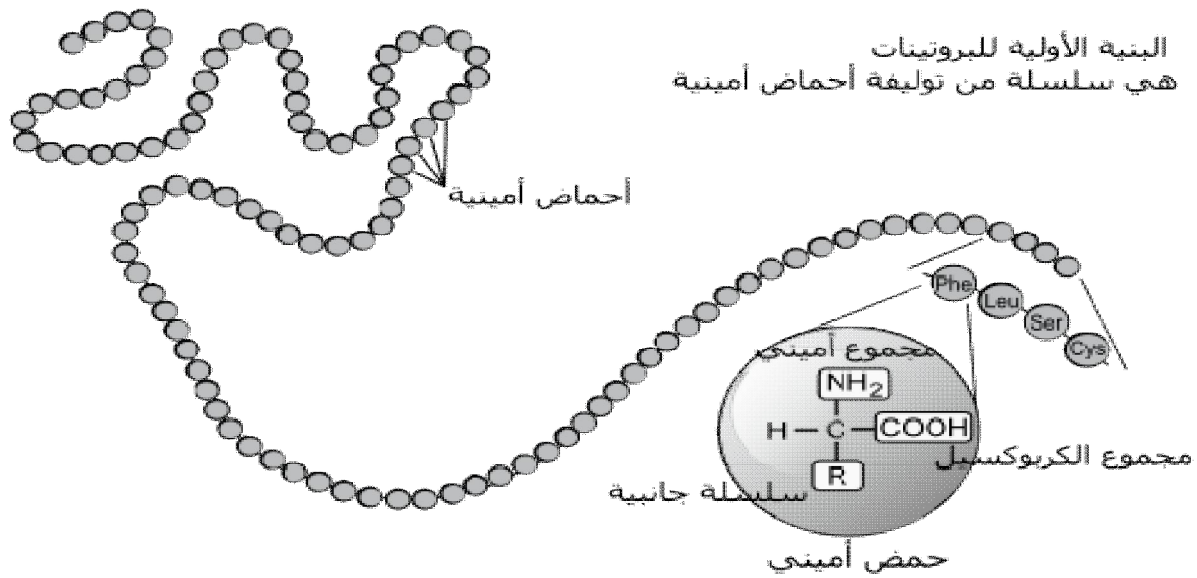
RNA	DNA	المقارنة
يتكون من سلسلة واحدة	يتكون من سلسلتين	عدد السلاسل
يصنع في النواة وينتقل إلى السيتوبلازم	يوجد في النواة	مكان وجوده في الخلية
سكر خماسي الكربون	سكر خماسي الكربون منقوص الأوكسجين	نوع السكر

الجدول 1.2 : يوضح الفرق بين RNA و DNA

## 4.3.1.1.2 البروتينات

يتركب البروتين من عدد من الأحماض الأمينية (مواد تحتوي على مجموعة كربوكسيل-COOH حمضية ومجموعة أمين-NH<sub>2</sub> قاعدية لذا فإن لها تأثير متعادل)، ترتبط تلك الأحماض مع بعضها بروابط كيميائية تعرف بالروابط الببتيدية وهي رابطة تساهمية كيميائية .

يتميز كل بروتين ببنية مختلفة عن البروتينات الأخرى، تدعى هذه البنية بالحالة الأصلية للبروتين وتحدد حسب ترتيب الأحماض الأمينية في عملية الترابط التي تشكل السلاسل البروتينية.



الشكل 3.2 : بنية البروتين

## 4.1.1.2 نبذة تاريخية عن المعلومات الحيوية

ظهر مصطلح المعلومات الحيوية في الأدبيات العلمية عام 1991، ولكن بداية استخدامه ترجع الى عام 1968 مع ظهور كتاب ( أطلس البروتينات الأساسية و تراكيبها ).

المعلومات الحيوية علم جديد جمع بشكل أساسي ما بين علم الحاسوب وعلم الأحياء ففي المعلومات الحيوية يتم استخدام علم الحاسوب للإجابة على الأسئلة التي يطرحها علماء الأحياء من أجل فهم طبيعة الكائنات الحية. ولا يقتصر هذا على علم الحاسوب وعلم الأحياء فقط بل تستخدم علوماً أخرى كالإحصاء والرياضيات وعلم إسترجاع البيانات .

وقد كان ظهور هذا العلم نتيجة لزيادة كمية البيانات البيولوجية بشكل هائل في مدة قصيرة، ولعل من أهم البيانات البيولوجية الجين البشري الذي يتكون من 3 بلايين رمز ويميز الإنسان عن باقي الكائنات الحية ويمكن التحدي الأكبر بالنسبة لعلماء الأحياء في فهم تلك البيانات الهائلة والتوصل إلى أسرار طبيعة الكائنات وبالتالي فإن وجود فجوة كبيرة بين توافر البيانات البيولوجية وبين فهم الإنسان لتلك البيانات دعا إلى وجود المعلومات الحيوية كمجال علمي يُعنى بتقليص تلك الفجوة.

قام الدكتور روبرت ليدلي في سنة 1965 بنشر مقال في مجلة علمية بعنوان أساسيات التفكير في التشخيص الطبي "Reasoning Foundations of Medical Diagnosis" حيث دعا إلى إستعمال التكنولوجيا الرقمية

في مجال الطب والأحياء، وفي سنة 1984 ساهم مع عالمة مارغريت دايهو في تأسيس قاعدة البيانات PIR والتي تحتوي على جميع سلاسل البروتين المعروفة في ذلك الوقت.

مع تطور تكنولوجيا الإتصال وظهور الانترنت تمكنت العديد من الجامعات ومراكز الأبحاث من الإتصال بالإنترنت مما جعل هذا النوع من قواعد البيانات متاحاً للعديد من فرق البحث وأصبح لقواعد البيانات دوراً مهماً في مختلف البحوث البيولوجية.

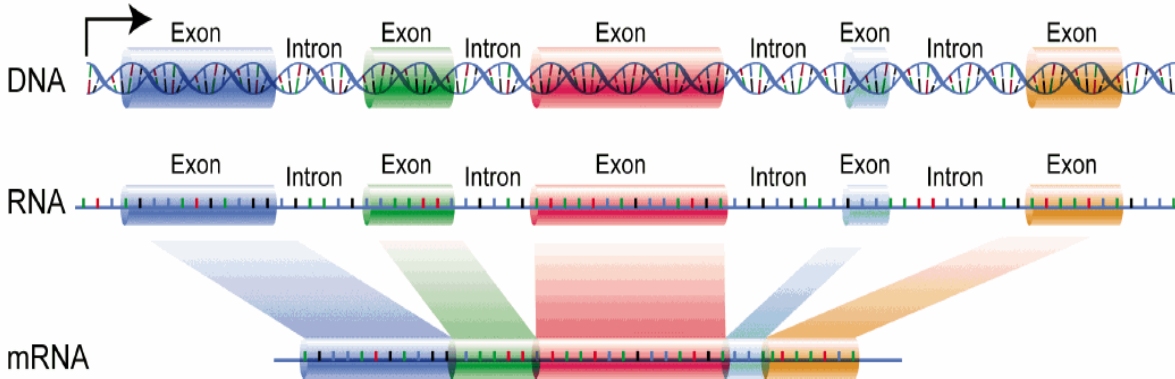
هذا التطور الجديد غيّر الكثير من المعطيات في مجال الأحياء، حيث سمح التوفر السهل لهذه المعلومات للباحثين بمقارنة تركيبية بروتين جديد مع بروتينات أخرى بطريقة أوتوماتيكية ومقارنة جينات بعض الفيروسات مع فيروسات أخرى مما أمكن من تصنيف العديد من البكتيريا والفيروسات وبالتالي إمكانية الحصول على فكرة أولية لكيفية معالجة فيروس جديد وإمكانية تحديده فصيلته.

ولضمان دقة المعلومات المخزنة في هذه القواعد أسست لجان مختصة في التنقيح والتدقيق فيها، فمثلاً إذا إكتشف فريق بحث بنية بروتين غير معروف أو تركيبية (DNA) لفيروس أو بكتيريا أو أي كائن حي آخر يمكنها إرسال هذه المعلومة إلى قاعدة البيانات المختصة حيث يتم مراجعتها ومراجعة الطريقة التي أجريت بها التجارب ومن ثم إضافتها.

ومع تضاعف حجم هذه البيانات بطريقة أسية بدأ أداء هذه القواعد في الترددي حيث أن عملية البحث أصبحت تستغرق وقتاً أكبر مما إستدعى تطوير خوارزميات فعالة للمقارنة بين مختلف السلاسل.

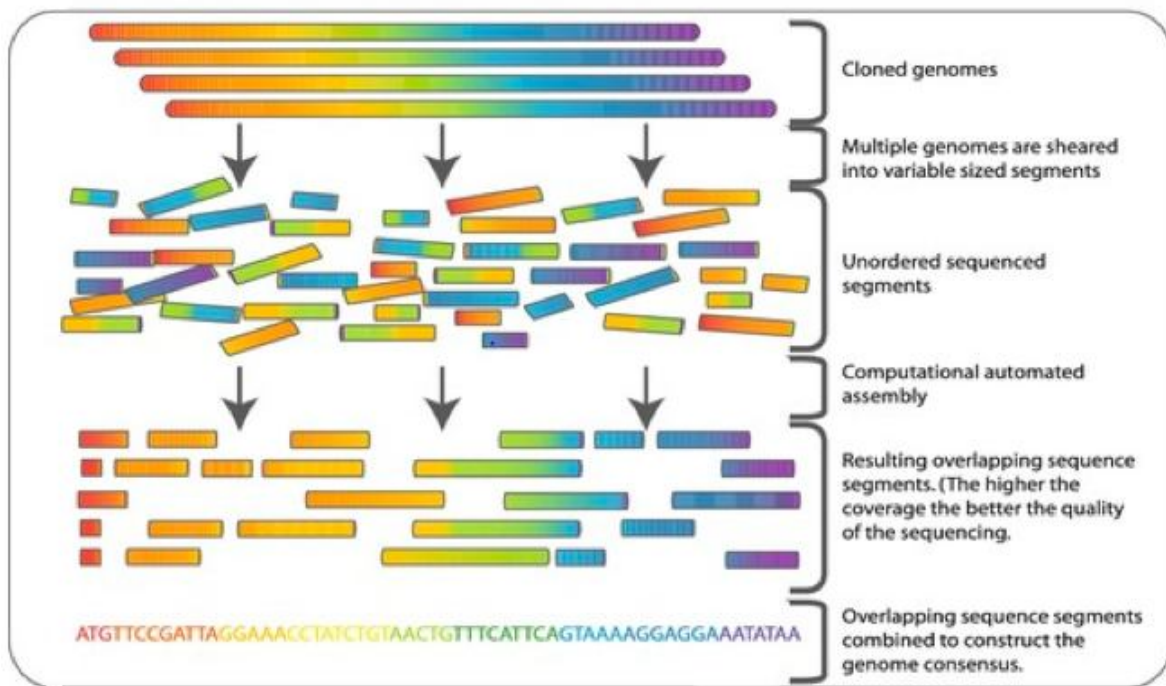
لكن هذا لم يكن كافياً فمع تطور تقنيات تشفير (DNA) وإستخلاص سلاسل البروتين أصبح بمقدور العلماء تشفير جينات تحتوي على ملايين النوكليوتيدات فأصبح من الضروري إيجاد طرق بحث فعالة أكثر للبحث عن سلسلة في وسط بحر من الرموز، فتم تطوير برنامج بلاست (BLAST) للبحث في ظرف لا يتعدى دقائق والبحث عن سلاسل مشابهة للسلسلة المراد البحث عنها في مختلف الكائنات، ويعتبر برنامج بلاست من البرامج الأكثر إستعمالاً من طرف علماء الأحياء إلى الآن وهو يعتبر من الأدوات الأساسية لكل طالب أحياء [3].

في بدايات القرن العشرين كان إهتمام علماء المعلومات الحيوية منصباً على كيفية تحسين طرق مقارنة السلاسل والشفرات الوراثية بطرق تكون ذات معنى بيولوجي أكثر وتطوير خوارزميات لتحديد ما إذا كانت شفرة وراثية ما تمثل جيناً أم لا و ماهي الأجزاء التي تكوّن هذا الجين.



الشكل 4.2 : رسم تمثيلي لمختلف أجزاء الجين

في سنة 1990 قام المعهد الوطني للصحة الأمريكي بإطلاق مشروع يتمثل في تشفير جينوم الإنسان، شارك في المشروع مئات من العلماء وبالإضافة إلى المخابر الأمريكية كان فيه عدة مخابر موزعة على 18 دولة نوع تقنية التشفير آنذاك كان يعتمد على طريقة تحليل الحمض النووي (DNA Analysis) المطورة والتي كانت بطيئة نوعاً ما ونظراً لكبر حجم الجينوم البشري حوالي 3.3 بليون زوج تم تقطيع (DNA) إلى أجزاء تحتوي على أجزاء مشتركة وتم توزيع كل جزء على مختبر ليقوم بتشفيره بشكل مشابه للعبة تركيب الصور المقطعة .



الشكل 5.2 : طريقة تحليل الحمض النووي (DNA)

تمكن العلماء من إكمال المشروع قبل الوقت المحدد وقاموا في سنة 2000 بنشر نسخة أولية عن (DNA) وتم نشر النسخة النهائية سنة 2003. لكن كان هنالك بعض التصحيحات الطفيفة وآخر تصحيح كان سنة 2009. ويمكن لأي إنسان أن يتصفح الجينوم البشري في موقع جامعة كاليفورنيا في سانتا كروز (UCSC) [4].

بعد نهاية المشروع تم إكتشاف عدة أمور: الأول هو أن فقط نسبة واحد بالمئة من (DNA) تقوم بدور في الخلية أو ما يسمى الجينات أما الباقي فهو فقط نفايات ، أما الثاني هو أن عدد الجينات المرزمة للبروتين أقل



بكثير مما كان يتوقع، حيث أن قدر العلماء في البداية وجود على الأقل 100,000 جين نظراً لتعقد بنية الإنسان لكن المفاجئة كانت أن جسم الإنسان يحتوي فقط على حوالي 20,000 إلى 25,000 جين وهو عدد أقل بكثير من عدد الجينات الموجودة في بعض الحشرات مما أثبت أن التعقيد البنيوي للكائنات لا يعكس بالضرورة تعقيداً على مستوى الخلية

## 5.1.1.2 أهداف المعلومات الحيوية

- استخدام البيانات لتحليل وتفسير النتائج بطريقة مجدية من الناحية البيولوجية .
- تطوير خوارزميات جديدة وتقنيات إحصائية تساعد في تحصيل المعلومات من مجموعات ضخمة من البيانات .

## 6.1.1.2 تطبيقات المعلومات الحيوية

تشمل تطبيقات المعلومات الحيوية الكثير من المجالات البحثية نذكر منها:

### 1.6.1.1.2 الطب

- إكتشاف العقاقير : إكتشاف هيكل البروتين ثلاثي الأبعاد أدى إلى إحداث تغيير جذري في صناعة العقاقير حيث أنه يؤثر في كل مرحلة من مراحل تصميم العقار.
- الطب التشخيصي : يقترح الرعاية الصحية التي تناسب كل مريض، هذا المجال يتيح اتباع نهج جديد في التشخيص، تطوير الدواء والعلاج.
- الطب الوقائي : الرعاية الوقائية للحماية من الأمراض والإصابات بدلاً عن العلاج، هذا المنهج يعمل تحسين مستوى الصحة العام وليس الفردي. مثال للوقاية غسل اليدين والتحصين.
- العلاج الجيني : يمكن العلاج بالجينات عن طريق إساقعة جينات أو إزالة الجينات الضارة أو بالتحكم فيها.

### 2.6.1.1.2 الزراعة

- مقاومة الحشرات : يمكن إستخدام الجينات في التحكم في الآفات الخطيرة فيمكن تخفيض المبالغ المستخدمة في المبيدات الحشرية وبالتالي زيادة نوعية المحاصيل الزراعية.
- تحسين نوعية الغذاء : نجح العلماء في تحويل الجينات إلى أرز لزيادة مستوى فايتمين A ، الحديد وغيرهم من المغذيات وهذا العمل يمكن أن يؤثر في الحد من حالات العمى وفقر الدم [5] .

## 7.1.1.2 المشاكل التي تواجه المعلومات الحيوية

من المشاكل التي تواجه المعلومات الحيوية :

- الحجم الكبير للمعلومات الحيوية يؤثر في سرعة العمليات التي تحدث فيها.
- كيفية تحويل الملاحظات البيولوجية إلى نماذج يستطيع الحاسوب أن يفهما .

- تتطلب معرفة جيدة بلغات البرمجة مثل : Perl , C++ , Java كما يجب عليك الإلمام بلغة HTML .
- قلة عدد المتخصصين في المعلومات الحيوية .

# المصفوفات الدقيقة

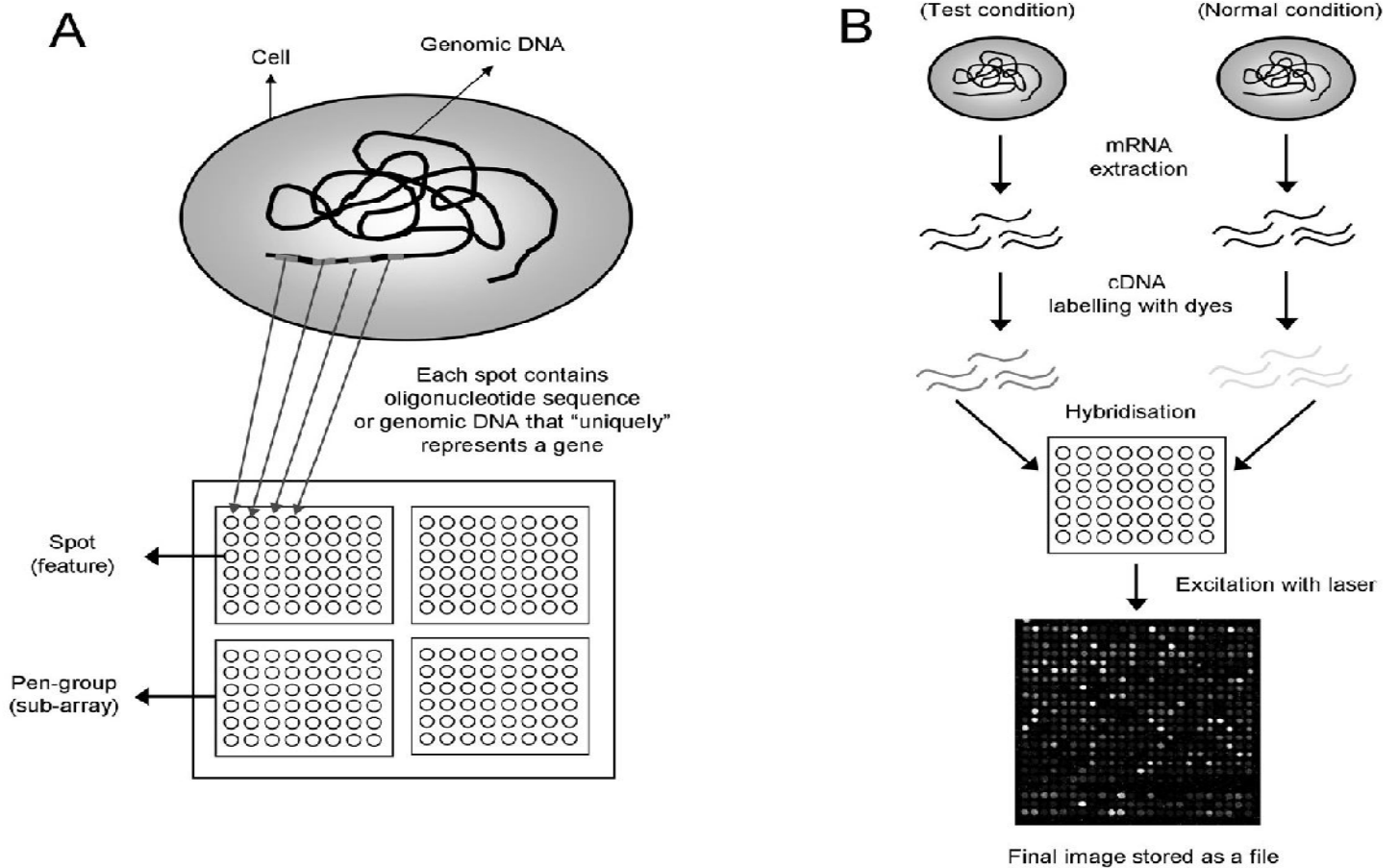
## 1.2.1.2 مقدمة

### 2.2.1.2 تعريف المصفوفات الدقيقة

هي عبارة عن تقنية جديدة تستخدم لفحص مادة حيوية معينة (DNA, protein , mRNA) باستخدام مستقبلات خاصة.

تتميز هذه التقنية بأنها توفر كمية هائلة من المعلومات باستخدام تجربة واحدة ، وعلى سبيل المثال : في حالة عمل اختبار لفحص الجينات في جسم الانسان فإن عدد الجينات تقريبا يساوي 30,000 جين لذلك في حالة فحص البروتينات الموجودة في عينة ما فإنه سيتطلب ذلك الكثير من الوقت والجهد ، ومن هنا جاءت هذه التقنية لتسمح للعلماء بأخذ النتائج في وقت قصير.

هذه التقنية تكون على شريحة زجاجية أو على خلية رقيقة مصنوعة من السيلكون أو البلاستيك المصنوعة الدقيقة مكونة من عدة مناطق وكل منطقة تسمى (feature) ، المستقبلات الخاصة تكون مرتبطة على الشريحة الزجاجية أو الخلية الدقيقة ودورها يكون في الإرتباط بالمادة المراد فحصها بطريقة أخرى هذه التقنية عبارة عن مجموعة من الإختبارات المخبرية وضعت في شريحة صغيرة [6].



الشكل 6.2 : مصفوفة إختبار

## 3.2.1.2 إستخدامات المصفوفات الدقيقة

يوجد عدة استخدامات لهذه التقنية ومنها :

### 1.3.2.1.2 مصفوفة البروتين الدقيقة (protein microarray)

يتم الاستفادة من هذا النوع في عملية فحص البروتينات من حيث وجودها أو عدمه وأيضاً يتم قياس مستوى إفراز البروتين في العينة ومن أشهر المواد المستخدمة للإرتباط بالبروتينات هي الاجسام المضادة (antibody) أما في حالة الكشف عن الأجسام المضادة فتستخدم المستضدات (antigen) كمواول للإرتباط أظهرت دراسة في عام 2010 أنه تم استخدام مصفوفة بروتين شملت على 810 جسم مضاد موجهه ضد 741 بروتين لها ارتباط بمرض السرطان [7] .

#### 1.1.3.2.1.2 تطبيقات مصفوفة البروتين

مصفوفة البروتين لها تطبيقات عدة منها :

- فحص الأجسام المضادة .
- فحص السيتوكين (cytokine) في حالة مرض السرطان .
- تحديد نشاط الإنزيمات .
- دراسة تفاعل البروتينات مع بعضها البعض .
- الكشف عن تأثير الادوية .
- الكشف عن مؤشرات حيوية (biomarker) جديدة للأمراض على سبيل المثال : إرتفاع مستوى أو ظهور بروتين معين في المراحل الأولى لمرض السرطان قد يكون محدد حيوي يتم الإعتماد عليه في التشخيص أو الفحص المبكر للمرض .

## 2.3.2.1.2 مصفوفة الحمض النووي الدقيقة

### :(DNA microarray)

من المعلوم أن (DNA) يتكون من شريطين متصلين بمنطقة تسمى سنترومير، وأن كلا الشريطين يعتبران مكملان لبعضهما البعض أي أنه في حالة وجود شريط واحد فقط فإنه بالإمكان معرفة ترتيب القواعد النتروجينية في الشريط الآخر وهذه الميزة هي ما تعتمد عليه تقنية مصفوفة الحمض النووي.

كل خلية في جسم الإنسان تحتوي على نفس المحتوى الوراثي من الجينات ولكن نشاط بعض الجينات لا يتم إلا في خلايا محددة، مثلاً الإنسولين لا يتم إفرازه إلا في خلايا البنكرياس بسبب أن الجينات يتم تحفيزها فقط في هذه الخلايا لذلك هذه التقنية لها القدرة على الكشف عن الجينات المفعلة .

## 3.3.2.1.2 مصفوفة النسيج الدقيقة (Tissue microarray)

نوع اخر من المصفوفات والتي تكون مخصصة لإجراء الإختبارات على الأنسجة، في البداية يتم تحضير قالب المصفوفة ومن ثم يتم أخذ خزعة من القالب (block) المراد إجراء التجربة عليه ووضعها في قالب المصفوفة بعد ذلك يتم تقطيع القالب وتوضع على شريحة زجاجية وتسمى هذه العملية بـ Arraymold .

### 1.3.3.2.1.2 إستخدامات مصفوفة النسيج

من إستخدامات مصفوفة النسيج :

- تستخدم للصبغ .
- فحص المؤشرات الحيوية (biomarker) لمرض السرطان .



## 4.2.1.2 أنظمة المصفوفات الدقيقة

### 1.4.2.1.2 نظام القناة الواحدة

في هذا النظام يتم ربط العينة بمادة حاملة للإشعاع (fluorophore) ومن ثم توضع على المصفوفة الدقيقة مثال عليها مصفوفة الحمض النووي (DNA affymetrix microarray) وتسمى أيضا بـ AffymetrixGeneChip حامل الإشعاع عادة ما يكون مادة biotin.

### 2.4.2.1.2 نظام القناتين

يتم استخدام نوعين من العينات ( عينة سليمة (control) و العينة المراد فحصها (sample) )، لذلك يوجد أيضاً نوعين من حاملات الإشعاع السيانين 5 (cy5) والذي يصدر اللون الأحمر ويتم وضعه على العينات المراد فحصها والسيانين 3 (cy3) ويصدر لوناً أخضر ويستعمل للعينة السليمة تستخدم للمقارنة بين العينتين وملاحظة الإختلافات في النهاية يتم التمييز بلون الإشعاع وهي كالتالي :-

- الأحمر ..... sample
- الأخضر ..... control
- البرتقالي ..... sample +control
- الأسود ..... none

عندما يصدر اللون البرتقالي، هذا يدل على تواجد المادة في العينتين كما أن اللون الاسود يدل أن العينتين لا تحتوي على المادة المراد الكشف عنها[7].

## 5.2.1.2 مشكلة المصفوفة الدقيقة

- صور المصفوفة الدقيقة تشمل آلاف البقع كل منها يقابل مادة حيوية مختلفة، إن مشكلة إنقسام المصفوفة الدقيقة تحسب كل نقطة في الشاشة كبقعة.
- تتطلب كميات كبيرة من البيانات للتدريب.

# الفصل الثاني

1.2.2 علم تعلم الآلة

2.2.2 التعرف على الأنماط

3.2.2 الشبكات العصبية

4.2.2 التعلم المتعمق

5.2.2 الدراسات السابقة



# تعلم الآلة

## 1.1.2.2 مقدمة

علم تعلم الآلة (Machine Learning) هو أحد فروع علم الحاسوب يتفرع من علوم الذكاء الصناعي، جزء منه يدرس منح الآلات و الحواسيب القدرة على التعلم دون أن يتم برمجتها صراحة من خلال إستكشاف خوارزميات تستطيع تعلمها ذاتياً وتصنع التنبؤات بخصوص البيانات، فهذه الخوارزميات تعمل من خلال بناء نماذج من المدخلات وذلك للمساعدة في إتخاذ القرارات هذا بدلاً من أن تتبع الآلة تعليمات وشفرات برمجية ثابتة ومقيدة لها يحددها البشر.

يتداخل علم تعلم الآلة مع علم إحصاء الحوسبة وهو فرع يركز أيضاً على صنع التنبؤات من خلال استخدام الحاسوب، كما أنه يرتبط بشدة مع علم التحسين الرياضي والذي يتركز إهتمامه على إختيار البديل الأفضل من بين عديد من البدائل المتاحة ويوفر لتعلم الآلة كثير من الوسائل والنظريات والتطبيقات [8].

## 2.1.2.2 خوارزميات التعلم

تنقسم أنواع المشكلات والمهام لتعلم الآلة إلى ثلاثة تصنيفات طبقاً لطبيعة إشارة التعلم أو التغذية المرتدة المتاحة لنظام التعلم كالاتي:

### 1.2.1.2.2 التعلم المراقب

هو أن يتم إعطاء أمثلة المدخلات للآلة والمخرجات المرغوبة من قبل المعلم -البشر - والهدف أن تتعلم الآلة قاعدة عامة تربط المدخلات بالمخرجات.  
من أهم هذه الأنواع :

#### 1.1.2.1.2.2 التصنيف (Classification)

وهو النوع الأكثر استخداماً في تعلم الآلة في هذا النوع يكون الدخل مصنفاً الى نوعين أو أكثر وهدف عملية التعلم إنتاج نموذج يستطيع تصنيف أي دخل جديد الى نوع أو أكثر من الأنواع المعرفة سابقاً ومن أمثلة هذا النوع عملية تصنيف البريد الإلكتروني وعملية التعرف على الوجوه .

#### 2.1.2.1.2.2 الإنحدار (Regression)

يقوم على التنبأ بقيم مستمرة بدلاً من أصناف منفصلة هنالك العديد من التطبيقات لهذا النوع كالتنبؤ بأسعار البورصة والتنبؤ بعمر شخص يشاهد مقطع من الفيديو والتنبؤ بدرجة الحرارة داخل مبنى اعتماداً على معلومات الطقس والوقت والحساسات الموجودة .

## 2.2.1.2.2 التعلم غير المراقب

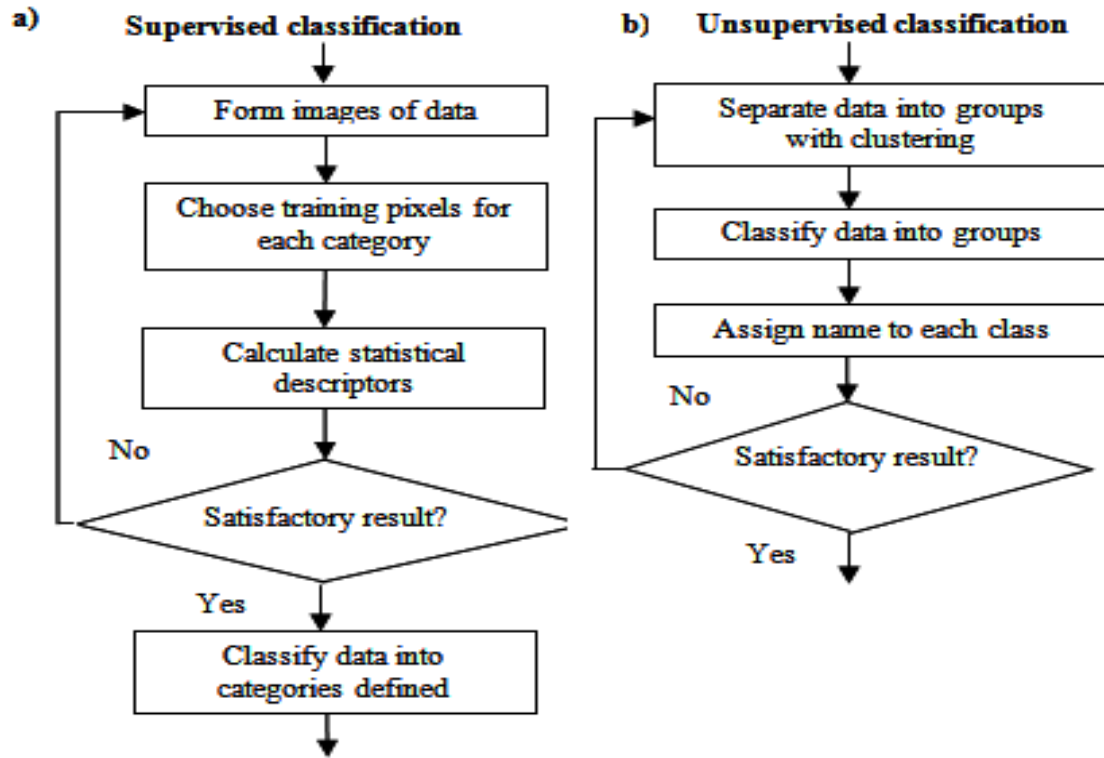
تترك خوارزمية التعلم لتعتمد على نفسها في إستكشاف هيكل مدخلاتها بهدف أن تكتشف الأنماط الخفية في البيانات و من أهم الأنواع الفرعية ضمن هذا النوع :

### 1.2.2.1.2.2 التجميع (Clustering)

في هذا النوع يتم فرز الدخلى الى مجموعات غير معروفة مسبقاً ومن تطبيقاته تعلم حركات الشخص الواقف أمام كاميرا تقوم بتسجيل حركاته بحيث يستطيع النظام لاحقاً التعرف على هذه الحركات وربطها بردود فعل مناسبة .  
و من التطبيقات الأخرى في مجال التجارة الإلكترونية عملية تجميع المستخدمين في مجموعات بناء على عملية الشراء التي قاموا بها و سلوك التصفح الخاص بهم، و من ثم إستخدامها لإرسال رسائل إعلانية موجهة بحسب كل مجموعة .

### 2.2.2.1.2.2 التعلم التعزيزي (Reinforcement Learning)

وهو الأقل إستخداماً ، وفي هذا النوع يتم تعلم كيفية التصرف عند حدث معين من خلال إعطاء إشارات ترمز الى مكافئة أو عقاب بناء على السلوك الحالي [9].



الشكل 7.2 : يوضح التعلم المراقب والتعلم الغير مراقب

## 3.1.2.2 أمثلة خوارزميات تعلم الآلة

- Linear Classifiers
- Logical Regression
- Naïve Bayes Classifier
- Perceptron
- Quadratic Classifiers
- K-Means Clustering
- Boosting
- Decision Tree
- Support Vector Machine
- Neural networks
- Bayesian Networks

## 4.1.2.2 تطبيقات تعلم الآلة

تعلم الآلة يُوظف عادة في نطاق واسع من مهام الحوسبة خاصة التي تكون الخوارزميات البرمجية الصريحة والمقيدة فيها غير مجددة عملياً ومن أمثلتها :-

- المعلومات الحيوية.
- معالجة اللغات الطبيعية .
- محركات البحث .
- مجال التنبؤات بالطقس.
- الحوسبة المالية .
- التسويق .
- التشخيص الطبي .
- رؤية الحاسوب .



# التعرف على الأنماط (Patterns Recognition)

## 1.2.2.2 مقدمة

النمط هو كيان غامض التعريف يمكن إطلاق اسم عليه، فقد يكون بصمة أصبع، كلمة مكتوبة، وجه إنسان، صوت، أو حتى شريط الحمض النووي DNA والتعرف على الأنماط هو إكتشاف الفئة أو الصنف (Class) الذي ينتمي إليه هذا النمط .

تهدف البحوث والتقنيات الخاصة بهذا العلم إلى تطوير تقنيات للتعرف على أنماط وهيكل محددة في الإشارات الرقمية، حيث يمكن للنمط أن يمثل صورة تحتوي على حرف مكتوب أو مقطع موسيقي أو مقطع كلامي يمثل كلمة أو نص حاسوبي. ويمكن أن يكون النمط المطلوب التعرف عليه هو الحرف الذي تحويه الصورة أو الآلة المستخدمة في المقطع الموسيقي أو الكلمة الملفوظة في المقطع الكلامي أو تحديد ما إذا كان النص الحاسوبي يتحدث عن الفيزياء أو الأدب أو السياسة.

## 2.2.2.2 مجالات التعرف على الأنماط

تهدف البحوث والتقنيات الخاصة بهذا العلم إلى إيجاد أو تطوير تقنيات للتعرف على أنماط أو هيكل محددة في الإشارات الرقمية :

- التعرف على الكتابة : التعرف على الأحرف في الصورة التي تحوي كتابة وتحويلها إلى نص حاسوبي .
- التعرف الضوئي على الحروف (OCR) :  
تقوم هذه التقنية بقراءة وتحويل المستندات المصورة والتي بها نصوص مكتوبة والمدخلة للكمبيوتر بالماصح الضوئي إلى نصوص حية يمكن تحريرها ونسخها ثم التعديل عليها و من أمثلتها آلات القراءة للمكفوفين .
- القياسات الحيوية :
  1. التعرف على الوجه .
  2. التعرف على بصمات الأصابع .
  3. التعرف على الكلام : التعرف على الكلام الملفوظ في نص صوتي وتحويله إلى نص حاسوبي .
  4. التعرف على المتكلم : معرفة من هو الشخص المتكلم في مقطع صوتي .

- التشخيص الطبي .
- التطبيقات العسكرية : التعرف الآلي على الهدف .
- التعرف على الصور الفضائية والجوية .

## 3.2.2.2 الخطوات العامة لنظام التعرف على الأنماط

يتكون نظام التعرف على الأنماط من أربع مراحل و فيما يلي توضيح لهذه المراحل :

### 1.3.2.2.2 المعالجة الأولية (Preprocessing)

في هذه المرحلة يتم تهيئة النمط للمراحل التالية وذلك بعمل بعض عمليات المعالجة بغرض التحسين من هذه المعالجات، ويتم إزالة التشويش (Noise remove) من النمط، وتبسيطه بهدف تقليل كمية البيانات المعالجة .

### 2.3.2.2.2 إستخلاص الخصائص المميزة (Features Selection)

في هذه المرحلة يتم إيجاد أفضل مجموعة من الصفات والخصائص من النمط التي تساعد في عملية التصنيف .

### 3.3.2.2.2 التصنيف (Classification)

ندخل الخصائص المميزة المماثلة في شكل مصفوفة تسمى (Features Vector) ومن ثم نحدد نوع النمط الذي ينتمي له ال Vector ومن ثم نستخدم التقنيات الموجودة في التصنيف ويتم التعلم عن طريق عرض مجموعة من الأنماط على المصنف وتتم هذه العملية إما بواسطة تعليم مراقب أو تعليم غير مراقب .

### 4.3.2.2.2 مرحلة ما بعد المعالجة (Post Processing)

نقوم بتحويل مخرج النظام إلى شكل يمكن من التعامل معه ومحاولة معالجة الأخطاء الملاحظة بهدف تحسين الدقة العامة للنظام .

## 4.2.2.2 طرق التعرف على الأنماط

هناك أربع طرق أساسية مستخدمة في التعرف على الأنماط :

### 1.4.2.2.2 طريقة مطابقة القوالب (Template-Matching)

هذه الطريقة تقوم على تخزين مجموعة من القوالب (Template) أو النماذج (Prototypes) من كل صنف في الحاسوب وفي مرحلة التصنيف تتم عملية مقارنة الصور الداخلة مع القوالب فإن كانت نتيجة مقارنتها مع الصنف S أكبر من نتيجة مقارنتها مع الصنف Y فإنها تصنف ضمن الصنف S وهكذا .

تتم عملية المقارنة بعد أن تخزن الصورة الداخلة على شكل مصفوفة وتُقارن مع القوالب الموجودة في الجهاز و تعطي قيمة للمقارنة الصعوبة في هذه الطريقة هي الإختيار الجيد للقوالب من كل صنف بالإضافة إلى تحديد معايير المقارنة وخصوصاً لو كانت الصورة الداخلة تحمل تشوهات .

### 2.4.2.2.2 الطريقة الإحصائية (Statistical Approach)

في هذه الطريقة توصف كل صورة أو نمط بواسطة مجموعة من الخصائص والتي من الممكن أن نعبر عنها بواسطة قيم حقيقة في مرحلة التعلم يقدم كل نمط كمتجه من الخصائص، عادة تتم عن طريق تقسيم مساحة الصورة إلى مناطق مجزأة، كل منطقة تقارن مع صنف فمثلاً لو كنا نريد التعرف على صورة تفاحة، فأولاً نحدد ما هي خصائص التفاحة التي نخزنها في مرحلة التعلم، على سبيل المثال : اللون، الوزن، الشكل، الدوران وكذلك يتم التعرف على التفاحة، تقسم الصورة الى أجزاء وكل جزء نقارن الخصائص الموجودة فيه مع خصائص الصنف المخزنة وهكذا، الصعوبة في هذه الطريقة هي في إختيار مجموعة من الخصائص لكل فئة وقواعد القرار في التعرف على النمط .

### 3.4.2.2.2 طريقة التركيب (Structural Approach)

في هذه الطريقة لا نكتفي فقط بالقيم الرقمية لخصائص كل صنف، ولكن نضيف عليها العلاقات والخصائص في كل صنف والتي تتيح لنا معلومات هيكلية ضرورية في التعرف على الأنماط .

عملية التصنيف في هذه الطريقة تتم من خلال تحليل التراكيب أو بمعنى آخر برنامج تعريب، أعلى نسبة مقارنة ناتجة من مقارنة الصورة المدخلة من كل شجرة تحدد الصنف الذي تنتمي إليه الصورة المدخلة.

تستخدم هذه الطريقة في التعرف على الأهداف أو الصواريخ وكذلك التعرف على الأحرف وغيرها

## 4.4.2.2.2 الشبكات العصبية (Neural Network Approach)

الشبكات العصبية الاصطناعية هي إحدى التقنيات الحديثة نسبياً في الحوسبة، وهي مستوحاة من طريقة عمل العقل البشري والجهاز العصبي المركزي. تتكون الشبكات العصبية من عدد كبير من وحدات المعالجة العصبية المتشابكة تشابكاً كبيراً فيما بينها بحيث تكون قادرة على معالجة أنواع معينة من المشاكل كما في الخلايا العصبية الحية، تحتاج الخلايا العصبية الاصطناعية إلى التدريب بحيث يتم ضبط الأوزان فيما بينها بعد عملية التدريب، يمكن اعتبار الشبكة العصبية (خبيرة) في فئة المعلومات التي تم تدريبها عليها.

## 5.2.2.2 تطبيقات التعرف على الأنماط

تهدف البحوث والتقنيات الخاصة بهذا العلم إلى إيجاد أو تطوير تقنيات للتعرف على أنماط أو هياكل محددة في الإشارات الرقمية، حيث يمكن للإشارة أن تمثل صورة تحوي حرف مكتوب أو مقطع موسيقي أو مقطع كلامي يمثل كلمة أو حتى نص حاسوبي.

# الشبكات العصبية

## 1.3.2.2 مقدمة

### 2.3.2.2 تعريف الشبكات العصبية

عرفت الشبكات العصبية بأنها محاولة رياضية برمجية لمحاكاة طريقة عمل المخ البشري، حيث أن العلماء قد اكتشفوا تقريبا طريقة عمل المخ البشري من حيث قابلية التعلم وقابلية التذكر والقدرة على تمييز الأشياء والقدرة على اتخاذ القرارات، والمخ يتكون من مليارات الخلايا العصبية المتشابكة فيما بينها بطريقة معقدة جداً عن طريق الزوائد العصبية لكل خلية، مما يشكل شبكة هائلة من الخلايا العصبية المرتبطة فيما بينها عن طريق هذه الزوائد [10]

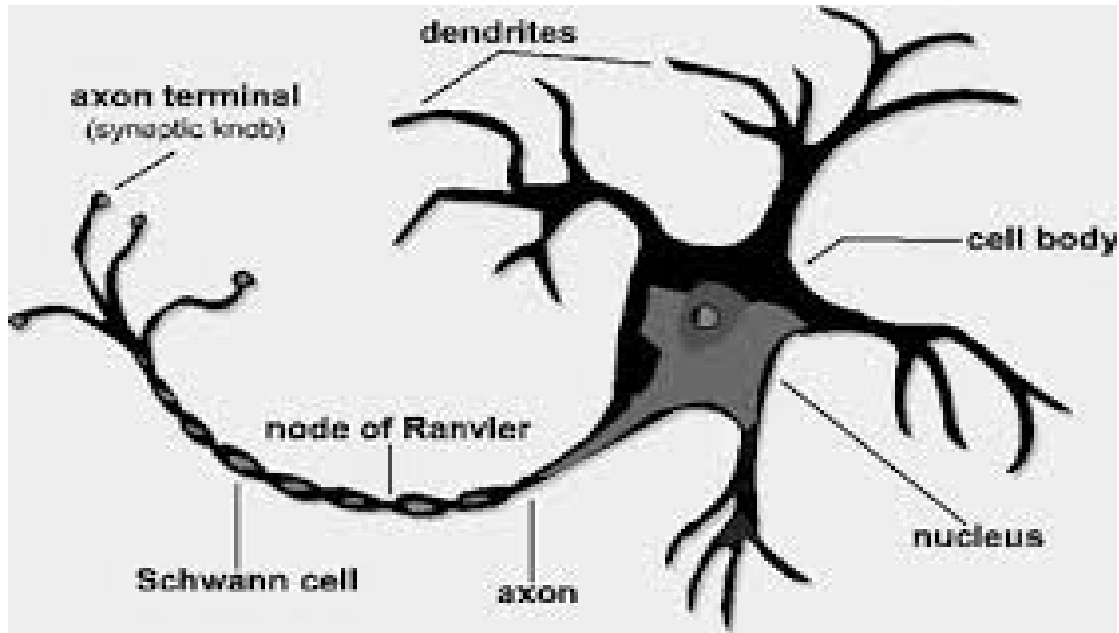
هذا الترابط بين الخلايا العصبية يتيح لها القدرة على تخزين المعلومات والصور والصوت وخلافه من الإشارات التي تصلها عبر الحواس الخمسة، ومن ثم تتيح لها أيضاً التعلم عن طريق التكرار والخطأ .

يكن أحد أسباب تفوق الدماغ البشري في قدرته على معالجة المعطيات في شكل هيكلية بمعنى أن كل جزء من المخ له وظيفة ويشرف على عمله جزء آخر يدير عمله ويستطيع أن يغير عمل الأجزاء ويسمى ذلك Plasticity كما أن كل جزء من المخ يعمل بصفة منفصلة عن الآخر كما لو كان جهاز كمبيوتر به عدد كبير جدا من المعالجات Processors، أجهزة الحاسوب اليوم تقوم بمحاكاة هذه العملية في ما يسمى حوسبة متوازية Computing Parallel، وبالرغم من السرعة العالية الناتجة عن هذه التقنية إلا أنها تفتقر إلى القدرة على الإستقلال بحل المشكلة، وذلك لأن الحاسب الإلكتروني يقوم بإرسال بيانات كثيرة جداً في صورة رقمية بين كل من أجزائه أما المخ فيقوم بعمل ذلك في صورة رقمية في بعض الأماكن وفي أماكن أخرى في صورة analog كما أن الكمبيوتر يرسل عدد كبير من البيانات المتكررة ولكن المخ لا يرسل بيانات متكررة ومثال على ذلك الكاميرا فهي ترسل حالة كل نقطة تصورها وإذا لم تتغير الحالة ترسلها كما هي أيضا أما العقل فيرسل التغيير فقط وذلك على شكل Spikes أو إشارات عصبية.

بشكل عام يمكن القول أن كل شبكة عصبية ترتب بشكل طبقات من الخلايا الاصطناعية : طبقة داخلية وطبقة خارجية وطبقة مخفية تتواجد بين طبقتي الدخل وطبقة الخرج .

وجد أن الخلايا العصبية تقوم بعملية جمع إشارات بمعنى أنه يوجد وصلتين لخلية عصبية مثلا كل وصلة عليها إشارة تكون النتيجة هي محصلة الإشارات بالجمع العادي ومن ثم وجد أن كل عصبون يستطيع أن يقوم

بعمل تكبير أو تصغير فتم إضافة عامل اسمه (Weighting Factor) بمعنى أنه إذا كان هناك خلية مثلاً ولها مدخلان فيتم ضرب الإشارة الأولى في المعامل الخاص بالعصبون الأول وضرب الإشارة الثانية في المعامل الخاص بالعصبون الثاني ومن ثم يتم جمعهم .



الشكل 8.2 : رسم توضيحي للشبكات العصبية الطبيعية ومكوناتها

### 3.3.2.2 خصائص الشبكات العصبية

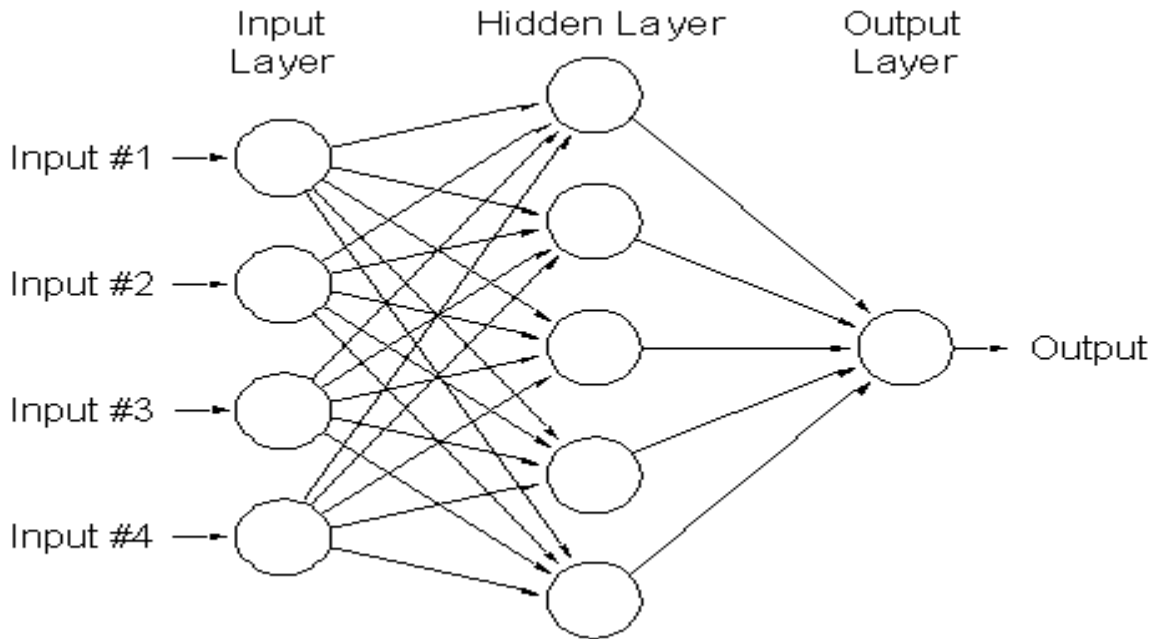
- 1- المرونة في التعامل مع أنظمة الصور المشوهة .
- 2- المقدرة على تمييز الصور المتفككة أو المتحللة جزئياً .
- 3- ونستطيع القول بأنها عبارة عن تركيبات ذات عمليات متوازية مع وجود عدد كبير من الوحدات العاملة التي تُحفز بواسطة ترابط العمليات الجارية وبالإضافة إلى تخزين من المعلومات الموزعة بصورة متوازية .
- 4- المقدرة على تقديم مخططات تتضمن علاقات غير خطية مضادة للضوضاء التي تجعل منها مصدراً جيداً لعمليات التصنيف والإسناد (classification predication).

## 4.3.2.2 طريقة معالجة المعلومات

كل إتصال بين عصبون وآخر يتميز بإرتباطه بقيمة تسمى الوزن (Weighting) وهي تشكل مدى أهمية الإرتباط بين هذين العنصرين، يقوم العصبون بضرب كل قيمة دخل واردة من عصبونات الطبقة السابقة بأوزان الإتصالات مع هذه العصبونات، من ثم جمع نواتج الضرب جميعاً ، الناتج يعتبر خرج العصبون الذي ينقل إلى عصبونات الطبقة اللاحقة .

## 5.3.2.2 نماذج الشبكات العصبية

الشكل النموذجي لهذه الشبكات هو ثلاث طبقات عصبونية على الأقل تدعى (طبقة دخل input layer ، طبقة مخفية hidden layer ، طبقة خرج output layer) طبقة الدخل لا تقوم بأي عملية معالجة فهي ببساطة مكان تغذية الشبكة بشعاع البيانات، تقوم طبقة الدخل بعد ذلك بتغذية (نقل المعلومات) الطبقة المخفية ومن ثم تقوم الطبقة المخفية بتغذية طبقة الخرج المعالجة الفعلية للبيانات تتم في الطبقة المخفية وطبقة الخرج أساساً .



الشكل 9.2 : يوضح شبكة عصبية إصطناعية بثلاث طبقات (طبقة دخل ، طبقة مخفية ، طبقة خرج)

## 6.3.2.2 أنواع الشبكات العصبية

أ. الشبكة العصبية أمامية التغذية هي إحدى أشهر أنواع الشبكات العصبية، ويحدث بها انتقال المعلومات عبر الطبقات باتجاه وحيد من طبقة الإدخال إلى الطبقة المخفية إلى الطبقة النهائية .  
ب. الشبكة العصبية خلفية التغذية هي إحدى طرق تعليم الشبكات العصبية التي تؤمن نقل معلومات بالانتشار العكسي للإتجاه الأصلي لقدم المعلومات.

تعتمد هذه الطريقة على مبدأ التعليم المراقب وتحتاج في مرحلة التدريب إلى بيانات خاصة تتعلم بها الشبكة حيث تقدم لها بيانات دخل مع بيانات الخرج المرغوب فيها ومن ثم تقوم الشبكة بعمل انتشار أمامي لبيانات الدخل للحصول على قيمة خرج الشبكة بعدها تقوم بالمقارنة بين الخرج المحسوب والخرج المرغوب فإذا لم تتطابق النتائج تقوم الشبكة بحساب قيمة الفرق بينهما لكل عصبون من طبقة الخرج والذي يمثل قيمة الخطأ، بعدها تأتي مرحلة الانتشار الخلفي للأخطاء حيث تعيد الشبكة حساب قيمة الخطأ في كل عصبون من الشبكات الخفية. في الأخير تأتي مرحلة تحديث قيمة الأوزان (weight update) حيث تقوم الشبكة بإعادة حساب كل الأوزان وتعويضها بالقيم الجديدة المحسوبة.

يشترط في الانتشار الخلفي أن تكون دوال التنشيط التي تستعملها العصبونات قابلة للإشتقاق. ذلك لأنه في مرحلة تحديث الأوزان تستعمل الدالة المشتقة لدالة التنشيط في حساب القيم الجديدة.

عندما يكون هناك عدد كاف من العصبونات، تكون الشبكة قادرة على التدرّب للقيام بأشياء مفيدة بالإستعانة بخوارزميات التدريب، تعتبر الشبكات أمامية التغذية مهمة جداً خاصة في استخدامات التصنيف والتمييز الذي لبيانات غير مألوفة مسبقاً [11] .

## 7.3.2.2 تطبيقات الشبكات العصبية

- تقريب الدوال .
- التعرف على الأشخاص .
- التعرف على المواقف .
- التعرف على الصوت أو الصورة.
- التعرف على الخطوط والكتابة باليد .
- التحكم .
- محاكاة الأنظمة .
- النمذجة .



# شبكات التعلم المتعمق

## 1.4.2.2 مقدمة

ظهر التعلم المتعمق منذ عام 2006 كمجال جديد ضمن بحوث تعلم الآلة، وقد تطورت تقنيات التعلم المتعمق البحثية خلال السنوات الماضية وأثرت على مجال واسع من العمل على معالجة الإشارة و المعلومات بشكلها التقليدي والحديث وذلك ضمن المجالات المتسعة المتضمنة المفاهيم الأساسية لتعلم الآلة والذكاء الاصطناعي.

## 2.4.2.2 تعريف شبكات التعلم المتعمق

- 1- صف من تقنيات تعلم الآلة التي تستخدم عدة طبقات من المعالجة غير الخطية للمعلومات لإستخراج ونقل الخواص بالتعلم المراقب أو غير المراقب بالإضافة لتحليل وتصنيف الأنماط .
- 2- حقل فرعي ضمن تعلم الآلة، يطبق خوارزميات التعلم على تمثيل متعدد المستويات، وذلك من أجل نمذجة العلاقات المعقدة ضمن المعطيات، وبذلك يتم تعريف السمات والمفاهيم العالية المستوى بناءاً على ما هو أدنى منها، تدعى هذه البنية الهرمية من السمات بالبنية العميقة، كما وتعتمد معظم هذه النماذج على منهجية التعلم غير المراقب للتمثيلات .
- 3- حقل فرعي ضمن تعلم الآلة يعتمد على تعلم عدة مستويات من التمثيلات المقابلة للبنية الهرمية من السمات أو العوامل أو المفاهيم، ويتم تعريف السمات والمفاهيم عالية المستوى بالإعتماد على المفاهيم ذات المستوى الأدنى، وقد تساعد مفاهيم المستوى الأدنى ذاتها في تعريف العديد من مفاهيم المستوى الأعلى . التعلم المتعمق هو جزء من عائلة أكبر من طرق تعلم الآلة المعتمدة على تعلم التمثيلات . إن مجال البحث الذي يهتم بإيجاد وتعريف هذه التمثيلات وكيفية تعليمها هو ما يطلق عليه التعليم المتعمق .
- 4- مجموعة من خوارزميات تعلم الآلة التي تحاول أن تتعلم في عدة مستويات لمثلها من التجريد، ويستعمل التعليم المتعمق الشبكات العصبية الاصطناعية وتتوافق المستويات بهذه النماذج التعليمية الإحصائية مع مستويات مختلفة من المفاهيم حيث يتم تعريف المفاهيم عالية المستوى بالإعتماد على المفاهيم ذات المستوى الأدنى وقد تساعد مفاهيم المستوى الأدنى ذاتها في تعريف العديد من مفاهيم المستوى الأعلى .
- 5- مجال جديد ضمن بحوث تعلم الآلة لتقريب تعلم الآلة بإتجاه واحد من أهدافه الأساسية وهو الذكاء الاصطناعي، ويدور حول تعليم مستويات متعددة من التمثيل والتجريد مما يساعد في تحقيق فهم أكبر للمعطيات كالصور، الفيديو والنصوص [12] .

ويعتمد التعليم المتعمق على مفهومين أساسيين مشتركين وهما :

- وجود نماذج مؤلفة من عدة مستويات من المعالجة غير الخطية للمعلومات .
- طرق تعلم تمثيل للسمات بإشراف أو بدون إشراف أكثر تجريباً في المستويات الأعلى .

يقع التعلم المتعمق في المنطقة المشتركة ما بين عدة مجالات وهي بحوث الشبكات العصبية والذكاء الاصطناعي والنمذجة الرسومية والأمثلية والتعرف على الأنماط ومعالجة الإشارات، وهناك ثلاثة أسباب وراء إنتشار التعلم المتعمق وهي :

1- القدرات المتزايدة لمعالجة الرقائق كوحدات المعالجة الرسومية عامة الأغراض

(General Purpose Graphical Processing Unit).

2- الحجم المتزايد للمعطيات المستخدمة في التدريب .

3- التقدم الحديث في تعلم الآلة ومعالجة الإشارة والمعلومات .

وقد سمحت هذه التطورات الأخيرة لوسائل التعلم المتعمق بإستعمال التوابع المعقدة المركبة غير الخطية من أجل تعلم التمثيلات الخواص الموزعة والهرمية . وتحقيق الإستعمال الفعال للمعطيات المصنفة وغير المصنفة، كما أثبتت دراسات العديد من الخبراء في عدة جامعات نجاح التعلم المتعمق في تطبيقات متنوعة كالرؤية الحاسوبية، التعرف اللفظي، البحث الصوتي، التعرف على الخطاب، ترميز الصفات الصورية والخطابية، التصنيف الدلالي للكلام وفهم اللغات الطبيعية، التعرف على خط اليد، معالجة الصوت، إستعادة المعلومات وتحليل الجزيئات التي قد تؤدي الى إكتشاف عقاقير جديدة[12].

## 3.4.2.2 تقنيات التعلم العميق

كانت معظم تقنيات تعلم الآلة ومعالجة الإشارة حتى وقت قريب تعتمد على معماريات ذات بنى سطحية ضئيلة حيث احتوت هذه المعماريات على مستويين من تحويلات الخصائص غير الخطية على الأكثر ومن أمثلتها :

• الأنظمة الديناميكية الخطية و غير الخطية .

• Gaussian Mixture Models (GMMs)

• Conditional Random Fields (CRFs)

• Maximum Entropy (MaxEnt)

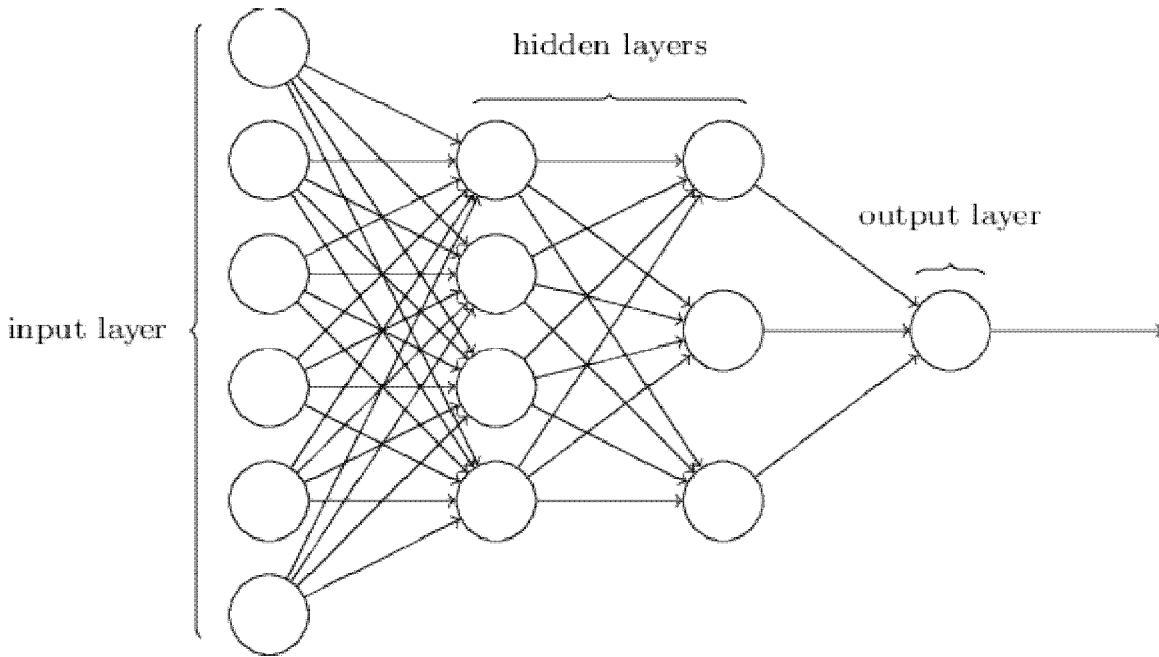
• Support Vector Machine (SVM)

• Logistic Regression

• Kernel Regression

وقد كانت هذه المعمارية السطحية فعالة في حل المشكلات البسيطة ولكن محدودة نمذجتها وقدرتها التمثيلية تسبب صعوبات في مواجهة التطبيقات الحقيقية الأكثر تعقيداً كالتطبيقات التي تحتوي على إشارات طبيعية كخطاب الإنسان والصوت واللغة الطبيعية والمشاهد البصرية .

ولكن ميكانيكية معالجة المعلومات ذات المصدر البشري تتطلب معماريات عميقة لإستخراج البنى المعقدة وبناء التمثيلات الداخلية بالإستفادة من الدخل المعطى .



الشكل 10.2 : يوضح شبكة تعلم متعمق (Deep Learning)

إستخدام الطبقات المخفية (Hidden Layers) ضمن الشبكة العصبية العميقة يحسن من قوة نمذجة هذه الشبكات وبدأ يشكل العديد من التشكيلات المثالية، فحتى وإن تعطل مسار ما ستستمر الشبكة العصبية العميقة الناتجة بأداء عملها بشكل جيد.

لكن استعمال الشبكات العميقة يتطلب طاقة حسابية كبيرة خلال عملية التدريب. وهذا سبب امتناع الباحثين عن العمل مع هذه الشبكات حتى فترة قريبة. كما ساهمت خوارزميات تعليم أفضل في نجاح الشبكات العصبية العميقة وكانت خوارزميات (SGD-Stochastic Gradient Descent) هي الأكثر فعالية مع

مجموعات التدريب الكبيرة كما أثبتت فعاليتها حديثاً في العمل التفرعي عبر عدة آلات في الوضع غير المتزامن، أو عبر عدة وحدات معالجة رسومية من خلال BP .

## 4.4.2.2 صنوف شبكات التعلم المتعمق

كما ذكرنا مسبقاً أن التعلم المتعمق يتعامل مع عدة طبقات من معالجة المعلومات غير الخطية والموجودة هرمياً بشكل طبيعي وبالاعتماد على طريقة استخدام البنى والتقنيات فمن الممكن تصنيف العمل في هذا المجال ضمن ثلاثة صنوف :

- 1- الشبكات العميقة للتعلم غير المراقب : تستخدم عند عدم توفر معلومات عن عدد صفوف الهدف أو تسمياتها حيث تقوم هذه الشبكات بإلتقاط الترابط العالي المستوى لمعطيات الدخل لمحاولة تحليلية والكشف عن أنماط هذه المعطيات .
- 2- الشبكات العميقة للتعلم المراقب : هي ممتازة في تصنيف الأنماط و ذلك عبر توصيف الصفوف الهدف بدقة وفقاً للمعطيات المرئية، و تكون البيانات الهدف المعرفة متوفرة دائماً سواء بشكل بشكل مباشر أو غير مباشر و تدعى الشبكات في هذه الحالة بالشبكات العميقة التمييزية .
- 3- الشبكات العميقة الهجينة (Hybrid Deep Network) : يكون هدفها تمييزي و يطبق على ناتج شبكات التعلم المتعمق المراقب أو غير المراقب وتكون الشبكات في هذا النوع أكثر جودة من شبكات النوع السابق [12] .

## 5.4.2.2 خوارزميات التعلم العميق

- Deep Belief Network
- Convolutional Neural Network
- Stacked Auto-Encoders
- Convolutional Auto-Encoders
- Feed forward Back propagation Neural Networks

# الدراسات السابقة

النتائج	التقنية المستخدمة	الباحث	اسم الدراسة
تمييز الأشخاص المصابين أو الغير مصابين بمرض التدرن الرئوي .	شبكة Hamming and Maxnet	محمد عبدالرحيم حمدي و كرم حاتم ذنون	تشخيص مرض التدرن الرئوي (السل) باستخدام الشبكات العصبية الاصطناعية
إستنتاج الممرات التنظيمية للخلايا والمساعدة في تشخيص الأمراض وتصميم الدواء المناسب وعمل خطة لمعالجة المرض . الدقة %83	Gene Expression Microarray (Gene chip)	Michael Molla, Michael Waddell, David Page, and Jude Shavlik	إستخدام تعلم الآلة لتصميم وترجمة تعبير جين المصفوفة الدقيقة
تشخيص المرحلة المبكرة من السرطان وتعريف نوعه عن طريق مقارنة نسيج متورم وآخر طبيعي. الدقة GSE8054 99% GSE3 94%	Supervised Learning <ul style="list-style-type: none"> <li>• k-nearest neighbors</li> <li>• SVMs</li> <li>• Boosting</li> </ul> Unsupervised learning <ul style="list-style-type: none"> <li>• k-means clustering</li> <li>• hierarchical clustering</li> </ul>	Pei-Chun Chen, Victoria Popic and Yuling Liu	تصنيف النسيج السرطاني باستخدام تعبير جين المصفوفة الدقيقة

تصميم نموذج لتمييز الصور والكتابة اليدوية. الدقة 95%	شبكة (Neocognitron Artificial Neural Network)	حنان حامد الدليمي و لهيب محمد الزبيدي	إستخدام الشبكات العصبية في التعرف على الأرقام العربية المكتوبة يدوياً
---	--	---	--

# الباب الثالث



# التقنيات المستخدمة

## 1.1 مقدمة :

كلمة ماتلاب هي اختصار لمصفوفة الإختبار (Matrix Laboratory) تمت كتابة الماتلاب في الأصل لتوفير إمكانية الوصول السهل إلى مصفوفة برمجة (Software Matrix) تم تطويرها من قبل مشروع LINPACK and EISPACK.

الماتلاب نظام تفاعلي عناصر بياناته تكون مصفوفة لا تتطلب الأبعاد، مما يسمح بحل المشاكل التقنية الحسابية بالأخص مشاكل المصفوفات وذلك في جزء من الوقت الذي يمكن أن يقضى في كتابة برنامج بلغة C أو FORTRAN .

يتميز الماتلاب بعائلة محددة التطبيق تسمى صناديق الأدوات، والتي تسمح بمعرفة وتطبيق التقنية المناسبة. تشمل صناديق الأدوات مجموعة دوال بإمتداد برنامج الماتلاب لحل مشاكل معينة تكون صناديق الأدوات متاحة في مجالات معالجة الإشارات، نظم المراقبة، منطق الغموض، الشبكات العصبية والمحاكاة وغيرها.

الماتلاب برنامج عالي الأداء لإجراء العمليات التقنية الحسابية، فهو يدمج البرمجة والحسابات والتصوير في بيئة سهلة الإستخدام حيث المشاكل والحلول تكون بصيغ رياضية [1].

## 1.2 خصائص الماتلاب:

- عالية المستوى للحوسبة العلمية والهندسية.
- استكشاف تكرارية التصميم وحل المشكلات.
- رسومات لتصوير البيانات وأدوات إنشاء المخططات.
- مهام محددة المجال لتصنيف البيانات، تحليل الإشارات وتطبيقات المنحنى.
- إضافة صناديق الأدوات على مجموعة واسعة من التطبيقات الهندسية والعلمية.
- أدوات إنشاء تطبيقات بواجهة مستخدم مخصصة (Custom User Interfaces) .

## 1.3 إستخدامات الماتلاب:

- الرياضيات والحساب.
- تطوير الخوارزميات.
- النمذجة والمحاكاة ووضع النموذج الأولي.
- تحليل البيانات والتقيب.

## 1.4 نظام الماتلاب :

يتكون نظام الماتلاب من خمسة أجزاء رئيسية :

### 1.4.1 لغة الماتلاب :

تعتبر لغة مصفوفة عالية الدقة مع تدفق التحكم، هياكل البيانات والمدخلات والمخرجات والبرمجة تسمح ببرمجة البرامج الصغيرة وكذلك التطبيقات المعقدة.

### 1.4.2 بيئة عمل الماتلاب :

مجموعة الأدوات التي تتعامل معها كمبرمج أو مستخدم للماتلاب، ويتضمن إدارة المتغيرات في مساحة العمل وإستيراد وتصدير البيانات، كذلك يتضمن أدوات لإعداد وتصحيح وإدارة Mat-files .

### 1.4.3 معالجة الرسومات :

تشمل أوامر عالية المستوى ثنائية وثلاثية الأبعاد لتصوير البيانات، معالجة الصور والرسوم المتحركة وعرض الصور. كما تتضمن أوامر منخفضة المستوى تتيح ظهور الرسومات كذلك بناء واجهات مستخدم رسومية في تطبيقات الماتلاب .

### 1.4.4 الدوال الحسابية :

تتضمن مجموعة كبيرة من خوارزميات الحوسبة مثل دوال الجمع والحسابات المعقدة إلى دوال المصفوفة وعكسها وغيره.

### 1.4.5 واجهة (API) :

تسمح بكتابة برامج بلغة C أو FORTRAN تتفاعل مع الماتلاب، تتضمن تجهيزات الإتصال الروتينية (Dynamic Linking) ويستخدم (Mat-files) للقراءة والكتابة [13].

# واجهة المستخدم الرسومية (GUI) :

واجهة المستخدم الرسومية تجعل البرامج أسهل إستخداماً من خلال تزويدها بمظهر متناسق مع أدوات تحكم أكثر سهولة مثل الأزرار والقوائم.

واجهة يجب أن تعمل بطريقة مفهومة ومتوقعة بحيث يدرك المستخدم ما عليه توقعه عند أي فعل يقوم به، على سبيل المثال عند الضغط بالماوس على الزر تقوم واجهة بالعمل الموضح على عنوان الزر المضغوط[14].

## طريقة عمل واجهة المستخدم الرسومية :

يتطلب وجود ثلاثة عناصر أساسية لعمل واجهة مستخدم رسومية للماتلاب :

**المكونات (Components):** كل قيد في هو مكون رسومي تشمل أنواع المكونات

عناصر تحكم رسومية مثل الأزرار والقوائم وغيرها، والعناصر الثابتة مثل الإطارات وتسلسل النص .

**الأشكال (Figures):** يجب أن يتم ترتيب مكونات واجهة الماتلاب في شكل والذي يعد نافذة في

شاشة الحاسوب . يمكن إنشاء أشكال فارغة بإستخدام دالة الأشكال والتي يمكن أن تستخدم في جمع المكونات.

**الإستجابة (Callback):** يجب على برنامج الماتلاب أن يستجيب لكل فعل إذا كان البرنامج

يتطلب أداء هذه الدالة، مثلاً إذا قام المستخدم بالضغط على زر يقوم الماتلاب بتطبيق دالة الكود لتنفيذ المطلوب.

## 1.5 مميزات الماتلاب :

- تمثل البيانات في الماتلاب في شكل مصفوفة، يعتبر العدد الصحيح مصفوفة من صف واحد وعمود واحد. جميع العمليات التي يمكن تطبيقها على المصفوفات تكون مضمنة في بيئة الماتلاب.
- إضافة مصفوفتان يتطلب عملية واحدة بدلاً عن for and while loop .
- يمكن إستخدام أدوات رسومية (Graphical Interactive Tools) لتغيير ألوان وأبعاد وأحجام البيانات.
- (Matlab Functionality) هي مجموعة من المهام تسمح بتطبيق دوال مميزة، مثلاً excel link تسمح بعرض البيانات في صورة Excel .

## 1.6 عيوب الماتلاب:

- يستخدم جزء كبير من الذاكرة، ويصعب استخدامه في الأجهزة البطيئة
- يأخذ وقت طويل من وحدة المعالجة المركزية (CPU) وهذا يجعل تطبيقات الوقت الفعلي معقدة (real-time applications) .

# الباب الرابع

# التطبيق العملي

## 1.4 المعالجة الأولية للبيانات (Preprocessing):

بيانات دراسة الحالة تم تنزيلها من موقع GEO وهو مستودع لبيانات الجينات مخزنة وموزعة على المصفوفات الدقيقة وأشكال أخرى من البيانات المقدمة من العلماء .

توجد عدة أشكال لتنزيل البيانات :-

-1 Full soft file

-2 Soft file

-3 Series family soft file

-4 Series family MINiML file

-5 Annotation soft file

The screenshot displays the GEO Dataset Browser interface. At the top, there is a search bar and navigation options. Below the search bar, a table lists 4348 DataSet records. The table has columns for DataSet, Title, Organism(s), Platform, Series, and Samples. The first few rows are:

DataSet	Title	Organism(s)	Platform	Series	Samples
GDS6248	Diet-induced obesity model: liver	<i>Mus musculus</i>	GPL6887	GSE39549	51
GDS6247	Diet-induced obesity model: white adipose tissue	<i>Mus musculus</i>	GPL6887	GSE39549	40
GDS6177	Acute alcohol consumption effect on whole blood (control group): time course	<i>Homo sapiens</i>	GPL570	GSE20489	25
GDS6176	Caspase-1 deficiency effect on lipid-loaded intestines	<i>Mus musculus</i>	GPL11533	GSE32515	18
GDS6100	MicroRNA-135b overexpression effect on prostate cancer cell line: time course	<i>Homo sapiens</i>	GPL10558	GSE57820	12
GDS6083	Chronic lymphocytic leukemia cells response to the neutralization of inhibitor of apoptosis proteins	<i>Homo sapiens</i>	GPL570	GSE62533	12
GDS6082	Sendai virus infection effect on monocytic cell line: dose response	<i>Homo sapiens</i>	GPL10558	GSE67198	11
GDS6064	Arthritic tarsal joints induced by collagen: time course	<i>Mus musculus</i>	GPL6246	GSE61140	15
GDS6063	Influenza A effect on plasmacytoid dendritic cells	<i>Homo sapiens</i>	GPL10558	GSE68849	10

Below the table, the detailed view for DataSet Record GDS6248 is shown. It includes tabs for Expression Profiles, Data Analysis Tools, and Sample Subsets. The Title is "Diet-induced obesity model: liver". The Summary states: "Analysis of livers of C57BL/6J mice fed a high fat diet for up to 24 weeks. Significant body weight gain was observed after 4 weeks. Results provide insight into the effect of high fat diets on metabolism in the liver." The Organism is *Mus musculus*. The Platform is GPL6887: Illumina MouseWG-6 v2.0 expression beadchip. The Citations are: Kwon EY, Shin SK, Cho YY, Jung UJ et al. Time-course microarrays reveal early activation of the immune transcriptome and adipokine dysregulation leads to fibrosis in visceral adipose deposits during diet-induced obesity. *BMC Genomics* 2012, Sep 4;13:450. PMID: 22947075. There is also a Cluster Analysis section with a heatmap and a Download button for the DataSet full SOFT file.

الشكل 1.4 : يوضح موقع GEO



تحتوي مجموعة البيانات على عدد معين من العينات، في هذا البحث تحتوي مجموعة البيانات على 42 عينة.

https://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS3716

Most Visited Getting Started Suggested Sites Web Slice Gallery Trashed

NCBI DATASET BROWSER GEO Gene Expression Omnibus

Search for GDS3716[ACCN] Search Clear Show All Advanced Search

DataSet Record GDS3716: Expression Profiles Data Analysis Tools Sample Subsets

**Title:** Breast cancer: histologically normal breast epithelium

**Summary:** Analysis of histological normal breast epithelia from both ER- and ER+ breast cancer patients and prophylactic mastectomy patients, and normal breast epithelia from reduction mammoplasty patients. Results provide insight into the mechanisms underlying breast cancer initiation and progression.

**Organism:** *Homo sapiens*

**Platform:** GPL96: [HG-U133A] Affymetrix Human Genome U133A Array

**Citation:** Graham K, de las Morenas A, Tripathi A, King C et al. Gene expression in histologically normal epithelium from breast cancer patients and from cancer-free prophylactic mastectomy patients shares a similar profile. *Br J Cancer* 2010 Apr 13;102(8):1284-93. PMID: 20197764

**Reference Series:** GSE20437 **Sample count:** 42

**Value type:** count **Series published:** 2010/06/01

Cluster Analysis

Download

- DataSet full SOFT file
- DataSet SOFT file
- Series family SOFT file
- Series family MINML file
- Annotation SOFT file

Data Analysis Tools

Find genes

Compare 2 sets of samples

Cluster heatmaps

Experiment design and value distribution

Find gene name or symbol:  Go

Find genes that are up/down for this condition(s):  specimen  disease state Go

الشكل 2.4 : يوضح عدد العينات (42 عينة)

ID_REF	IDENTIFIER	GSM51253	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	GSM51254	
1007_s_at	DDR1	2461.4	3435.7	1932.5	2377.7	3055.3	2978.1	2348.5	2963.9	2776.9	3088.9	3033.3	3037.1	3545.8	3322.6	1963.7	3609.6	2078.9	413.5									
1053_at	RFC2	26.7	159	31.2	140.7	69.9	98.5	37	59.9	86.7	107.2	64	82.9	97.7	69.7	82	45.6	84.5										
117_at	HSPA6	82.6	243.4	150.2	95.1	209.3	103.4	91.2	168.4	162.7	203.2	143.7	113.5	80	186.4	106.6	145.6	144.4										
121_at	PAX8	942.3	897.5	840.8	870.9	685.4	791.8	886.5	954.2	843.1	775.3	847.6	912.2	911.6	862.4	705	984.6	853.8										
1255_g_at	GUCA1A	71.8	87.9	75.4	58.1	31.8	40.3	70.5	43.3	51.6	42.6	74.9	53.7	30.5	15.2	42.5	76.6	88.2										
1294_at	UBA7	630.2	571.4	346.3	679.9	1289.3	421.1	417.6	811.6	778.1	393.2	995.4	987.7	938.5	924.6	480.8	1054.1	632										
1316_at	THRA	186	208.7	141	135	167.5	48.7	155	166.4	176.3	241.3	228.1	160	148.2	268.4	75.9	286.6	139.4										
1320_at	PTPN21	15.8	18	10.6	17.9	51.8	95.4	52.4	10.2	11.6	20.5	13.6	44.8	15.5	8.2	14.6	23.8	57.2										
1405_i_at	CCL5	71.3	26.1	26.3	23.8	100.1	10.9	4.6	65	12.7	59.3	23.3	58.8	67.1	26.1	9.8	116.4	22										
1431_at	CYP2E1	58.7	45	98.3	64.9	56.9	53	61	11.8	65.1	84.5	94.2	74.3	102.2	127.2	33	54	111.8										
1438_at	EPHB3	120.8	23.2	113.4	121	150.2	107.6	41.5	27.4	332.3	96.5	25.1	143.8	97.6	112.4	59.2	24.5	29.3										
1487_at	ESRRA	343.1	396.5	190.3	260	439.6	200.1	192.4	283.8	361.5	355.4	453	415.1	339.1	566.6	214.2	443	349.5										
1494_f_at	CYP2A6	327.7	438.4	310	267.7	248.8	279.6	257.4	304.3	242.5	474.2	301.4	256.2	188.2	292.4	201.9	454.2	333.5										
1598_g_at	GAS6	949.3	1143.8	940.2	1539.9	2623.8	853.5	949.9	1587.8	3006.2	1903.2	3380.7	2264.6	1499.8	3246.1	708.1	3137.1	1395										
160020_at	MMP14	540	467.9	430.5	424.3	527	214.1	311.2	251.5	504.7	519.1	419.2	386.3	267	435.9	370	500.4	316										
1729_at	TRADD	374.3	539.4	301.8	387.1	542.5	241.8	242.2	291.5	438.4	334.9	538.2	620.2	625.9	546.6	205.2	757.4	321.2										
1773_at	FNTB	91.5	52.9	83.3	41.5	30.3	80.6	66.5	49.1	75.8	60.7	18.6	19.1	106	55.4	70.5	24.6	123.4										
177_at	PLD1	142.1	170.2	180.3	222.4	135.2	139.7	90.6	154.5	110.2	218.6	118	146.2	83.6	154.7	118.2	103.2	117.7										
179_at	PMS2L11	285.4	444.7	254.9	352.6	378.3	387.6	274.7	291.4	435	397.6	490.6	446.5	458.4	631.4	300.1	765.4	309.6										
1861_at	BAD	5.3	52	5.2	68.6	77	43.2	16.4	27	35.4	13.2	88.2	90.8	110.9	134	21.7	113.3	32.7										
200000_s	PRPF8	1195.2	1546.7	928.1	1509.5	2694.3	1403.9	1415.7	1901.9	2119.7	1663.1	1396.2	1833.5	2178.7	1625.8	904.3	1950.6	1311										
200001_at	CAPNS1	724.8	372	587.5	842.2	1422.3	854.1	1133.9	1252.9	1146.1	712.6	892.7	912.7	1425.8	801.1	687.5	1511.9	998.5										
200002_at	RPL35	7788.5	7850.1	4413.7	3852.2	7130	5790.4	4705.3	7322.1	5244.5	4362.2	7447.3	6859.1	6526.7	10498.4	3998.4	9969.6	4652.3										

### الشكل 3.4 : يوضح مجموعة البيانات في ملف Excel

ما يهم حقاً في هذه البيانات هو معرفة الجينات وقيمها ( Value , gene IDs ) لتكون البيانات جاهزة للتحميل في الماتلاب على هيئة ملف (Excel Sheet).

يحتوي ملف Excel على أكثر من 22,000 جين قيمها موزعة على 42 عينة وكذلك يوضح حالة العينة إما (c) Controlled أو (b) with Breast Cancer.

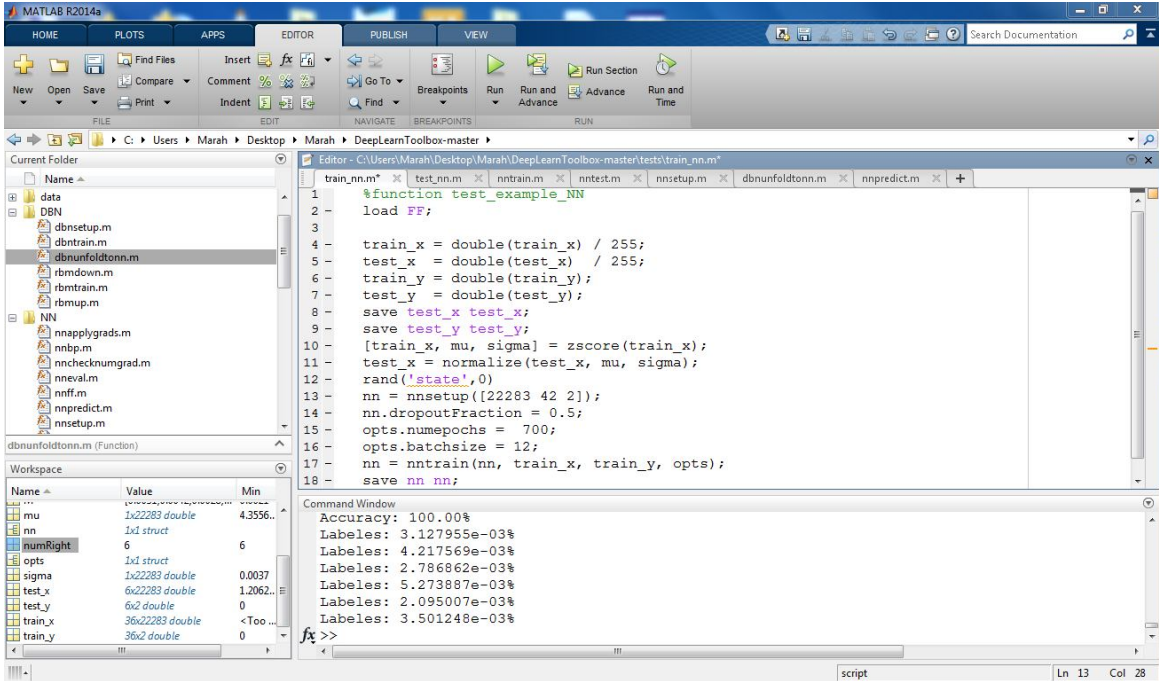
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
10	GSM512548	c	3088.9	107.2	203.2	775.3	42.6	393.2	241.3	20.5	59.3	84.5	96.5	355.4	474.2	1903.2	519.1	334.9	60.7
11	GSM512549	c	3033.3	64	143.7	847.6	74.9	995.4	228.1	13.6	23.3	94.2	25.1	453	301.4	3380.7	419.2	538.2	18.6
12	GSM512550	c	3037.1	82.9	113.5	912.2	53.7	987.7	160	44.8	58.8	74.3	143.8	415.1	256.2	2264.6	386.3	620.2	19.1
13	GSM512551	c	3545.8	97.7	80	911.6	30.5	938.5	148.2	15.5	67.1	102.2	97.6	339.1	188.2	1499.8	267	625.9	106
14	GSM512552	c	3322.6	69.7	186.4	862.4	15.2	924.6	268.4	8.2	26.1	127.2	112.4	566.6	292.4	3246.1	435.9	546.6	55.4
15	GSM512553	c	1963.7	82	106.6	705	42.5	480.8	75.9	14.6	9.8	33	59.2	214.2	201.9	708.1	370	205.2	70.5
16	GSM512554	c	3609.6	45.6	145.6	984.6	76.6	1054.1	286.6	23.8	116.4	54	24.5	443	454.2	3137.1	500.4	757.4	24.6
17	GSM512555	c	2078.9	84.5	144.4	853.8	88.2	632	139.4	57.2	22	111.8	29.3	349.5	333.5	1395	316	321.2	123.4
18	GSM512556	c	4138.6	31.7	133.6	846.8	90.6	448	494.1	16.8	15.2	87.1	164.7	898.9	263.5	9246.5	340.1	794.9	90.8
19	GSM512575	c	4520.4	54.7	132.7	718.6	50.2	1179.4	149.4	20.7	75.1	77.4	68.9	432.3	405.9	3093.6	396.3	719.4	43.6
20	GSM512576	c	3596.1	56.7	124.3	988.4	60	668.3	171.1	8.2	7.7	65.4	57.8	442.3	737.2	2464	393	498.8	27.2
21	GSM512577	c	2989	89.9	210.5	295.9	34.3	863.2	122.2	8.6	46.8	82.8	187.3	339.6	178.7	2728.6	156	426.6	72.9
22	GSM512578	c	3164.5	63.4	131.4	957.3	33.5	1055.5	159.4	7.7	120.3	54.7	16.7	377.2	255.2	4136.2	361.4	517.6	22.8
23	GSM512579	c	2764.3	57	89.6	630.8	61.7	1287.6	177.6	9.9	74	115.5	45.4	390.9	191.3	3740.9	253.2	444.9	133.8
24	GSM512580	c	4258.5	59.5	123.3	869.2	50.4	1127.8	245.4	14.4	79.8	69.2	100.2	467.5	277.2	3625.8	500.2	810.3	7.5
25	GSM512566	b	2836.6	77.1	120.9	425.7	59.2	1268.4	227.8	44.2	193.4	67.8	141.7	481	253.1	4164.5	394.4	668	148.6
26	GSM512567	b	2915.4	47.1	143.4	643.8	62.2	955.8	205.4	13.9	287.9	76.3	34.4	430.2	212.1	2610.3	197	608	141.4
27	GSM512568	b	3457.5	47	92.5	771.3	28.3	1157.5	206.3	10.5	48.9	59	13.5	291.7	239.2	6239.6	357.3	706.3	53.2
28	GSM512569	b	2798.7	83.2	72.1	681.1	97.6	888.6	153.4	8.6	7.6	98.3	53.5	442.4	514.8	1644.9	421.7	709.6	102.2
29	GSM512570	b	4370.2	40.2	131.8	812.7	8.1	1130.8	201.2	5	291.1	80.7	73.7	480.4	278.5	3110.9	546.8	901.5	24
30	GSM512571	b	2467.3	80.3	156.4	533.4	17.9	905.1	233.5	15.9	11	127.4	29.2	517.5	213.9	1700.6	201.1	441.6	70.6
31	GSM512572	b	3669.5	24.1	165.8	746.9	53	1138.5	196.7	49.4	69.8	67.2	73.7	424.8	205.3	7669.8	441.3	913.1	72.1
32	GSM512573	b	3310.1	8.8	141.6	1090.3	39.9	483	133.1	20.3	224.2	9.3	94.2	377.3	349.9	2445.8	574.8	506	76.6
33	GSM512574	b	3942.2	44.6	97.1	1008.7	11	1326.5	192.7	13.2	27.3	72.4	29.9	421.3	294	3366.6	562.7	531.2	13.3
34	GSM512557	b	4260.7	37.4	278.6	1273	65.8	1345.2	191	21.4	273.4	91.8	136.9	437.4	406.7	2326	490.1	767.3	43.7

الشكل 4.4 : يوضح حالة العينة (b , c)

بهذا الشكل تصبح البيانات جاهزة لتحميلها على الماتلاب .

## 2.4 تحميل البيانات :

يتم تقسيم البيانات إلى مصفوفتين ليتم إدخالها على شبكة DBN، الأولى تمثل المدخلات ( $train\_x$ ,  $test\_x$ ) والثانية ال Class ( $train\_y, test\_y$ ).



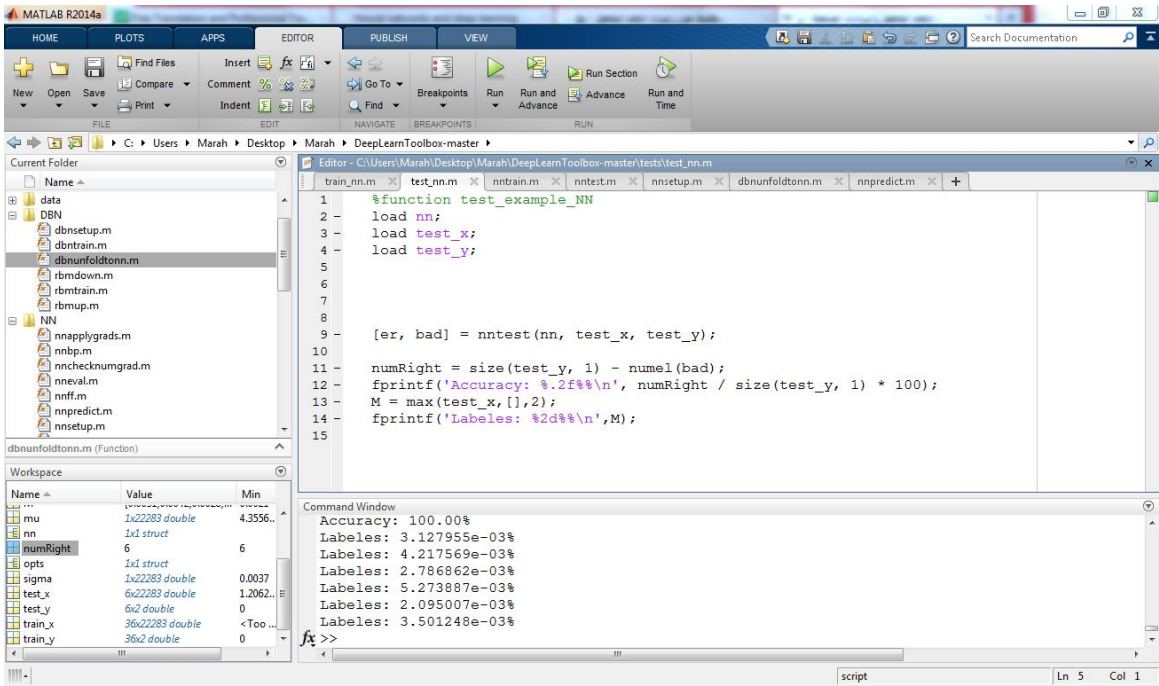
```
1 function test_example_NN
2 load FF;
3
4 train_x = double(train_x) / 255;
5 test_x = double(test_x) / 255;
6 train_y = double(train_y);
7 test_y = double(test_y);
8 save test_x test_y;
9 save test_y test_y;
10 [train_x, mu, sigma] = zscore(train_x);
11 test_x = normalize(test_x, mu, sigma);
12 rand('state',0)
13 nn = nnsetup([22283 42 2]);
14 nn.dropoutFraction = 0.5;
15 opts.numepochs = 700;
16 opts.batchsize = 12;
17 nn = ntrain(nn, train_x, train_y, opts);
18 save nn nn;
```

Name	Value	Min
mu	1x22283 double	4.3556e-03
nn	1x1 struct	
numRight	6	6
opts	1x1 struct	
sigma	1x22283 double	0.0037
test_x	6x22283 double	1.2062e-03
test_y	6x2 double	0
train_x	36x22283 double	<Too...
train_y	36x2 double	0

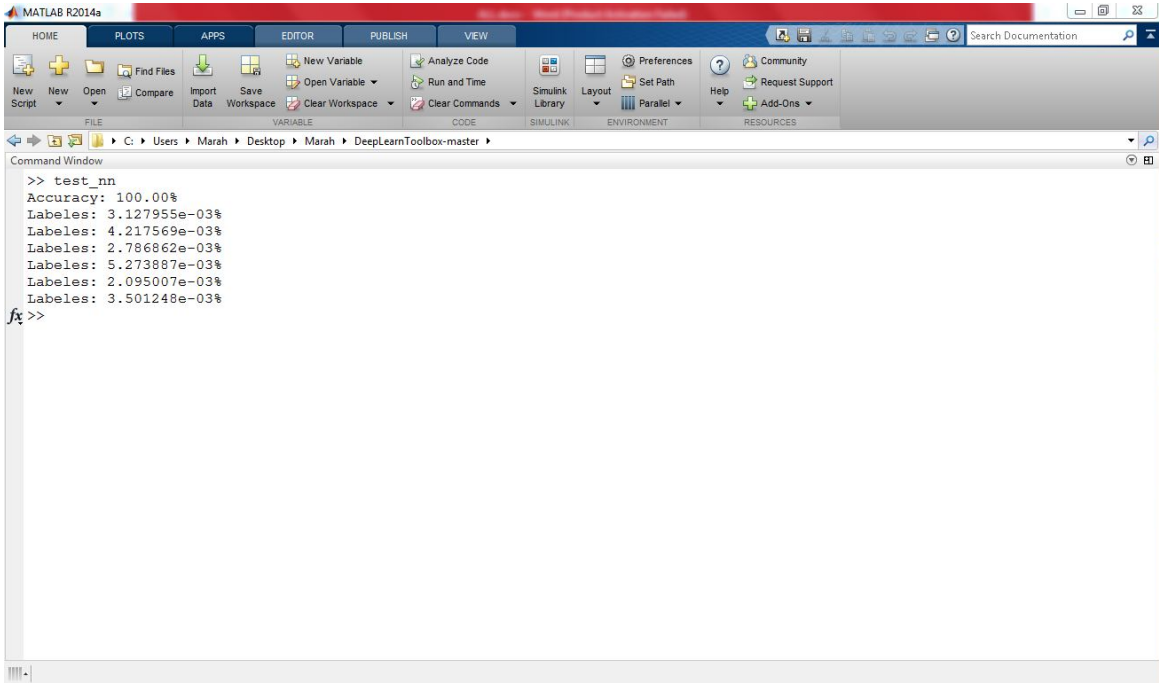
Command Window

```
Accuracy: 100.00%
Labels: 3.127955e-03%
Labels: 4.217569e-03%
Labels: 2.786862e-03%
Labels: 5.273887e-03%
Labels: 2.095007e-03%
Labels: 3.501248e-03%
```

الشكل 5.4 : يوضح شفرة التدريب



الشكل 6.4 : يوضح شفرة الإختبار



The image shows the MATLAB R2014a Command Window. The window title is "MATLAB R2014a". The Command Window contains the following text:

```
>> test_nn  
Accuracy: 100.00%  
Labels: 3.127955e-03%  
Labels: 4.217569e-03%  
Labels: 2.786862e-03%  
Labels: 5.273887e-03%  
Labels: 2.095007e-03%  
Labels: 3.501248e-03%  
fx >>
```

الشكل 7.4 : يوضح مخرجات الشبكة وأفضل خصائص

# الباب الخامس

# النتائج و التوصيات



## 1.5 المقدم

يتناول هذا الباب النتائج التي تم التوصل إليها بعد عملية تدريب وإختبار البيانات .

## 2.5 النتائج

No_epoch	Accuracy
10	20%
50	60%
120	60%
500	60%
700	60%
700*	100%

\*حصلنا على نسبة دقة 100% ب عدد مرات تدريب (EPOCH) = 700 بعد عملية Cross Validation .

نتيجة تدريب وإختبار شبكة NN نحصل على مخرجات تمثل أفضل خصائص لتشخيص مرض سرطان الثدي .

## 3.5 التوصيات

بعد الإنتهاء من هذا المشروع وتطبيقه وإتمامه نوصي بالآتي :

1. تطبيق هذا البحث في بيانات المصفوفة الدقيقة في السودان لتساعد في تشخيص ومعالجة الأمراض.
2. التعاون بين باحثين علوم الحاسوب والعلوم الطبية.
3. إستخدام نموذج آخر غير شبكة التعلم المتعمق لتصنيف البيانات.
4. هنالك خوارزميات مختلفة لتعلم المتعمق يمكن إستخدامها لتصنيف البيانات لمساعدة الباحثين في المجالات المختلفة.

## 4.5 الخاتمة

تم بحمد الله إتمام البحث والتوصل إلى النتائج السابقة، سائلين المولى عز وجل أن يجعل هذا العمل في ميزان حسناتنا، وما كان توفيق وسداد فهو من الله سبحانه وتعالى وأما النقص والزلل فهو من الشيطان، والحمد لله أولاً وأخيراً .

# المراجع

- [1] Cynthia Gibas, Per Jambeck, Developing Bioinformatics Computer Skills, April 2001.
- [2][www.mawdoo3.com](http://www.mawdoo3.com) , 2015-12-14 ، إسلام الزبون
- [3] جامعة مانشستر – بريطانيا ، صحيفة الحياة اللندنية ، 2007-5-29
- [4] إيهاب عبدالرحيم ، مجلة العربي ، يناير 2001
- [5] Madhu Bala Priyadarshi , Application of Bioinformatics , 18-9-2014 .
- [6] Dov Stekel , Microarray Bioinformatics , 2003
- [7] الحسن الرفاعي ، مختبرات العرب ، يناير 2012
- [8] Taiwo Oladipupo Ayodele , Introduction to Machine Learning, University of Portsmouth United Kingdom, February 2010.
- [9] Asoke Nath and Ariruna Dasgupta, Classification of Machine Learning Algorithms, International Journal of Innovative Research in Advanced Engineering (IJIRAE), March 2016.
- [10] David Kriesel , A brief introduction to Neural Network, 2007.
- [11] Simon Haykin, FEEDFORWARD NEURAL NETWORKS : AN INTRODUCTION , 1998 .
- [12] Li Deng and Dong Yu, Deep Learning: Methods and Applications, 2014.
- [13] <http://cimss.ssec.wisc.edu/wxwise/class/aos340/spr00/whatismatlab>
- [14] Introduction to Graphical User Interface (GUI) MATLAB 6.5 .