BREAST CANCER DETECTION USING ARTIFICIAL INTELLIGENCE TECHNIQUES

الكشف عن سرطان الثدي باستخدام تقنيات الذكاء الاصطناعي

Noor Hasan Mohsin SHUAIBT

نور حسن محسن شعيبت

Researcher ("Master's in Electrical and Computer Engineering") -Türkiye

باحثه (ماجستير في الهندسة الكهربائية والحاسوبية) في تركيا

Scientific supervisor

المشرف العلمي

Ali Ibrahim Lawah

علي إبراهيم لواح

Ph.D., "Information Security"

Ministry of Construction, Housing, Municipalities and Public Works- Iraq

أستاذ مشارك وزارة الاعمار والاسكان والبلديات والاشغال العامة- العراق

Abstract: *Cancer is one of the most dangerous diseases to humans, and yet no permanent cure has been developed for it. Breast cancer is one of the most common cancer types. According to the National Breast Cancer foundation, in 2020 alone, more than 276,000 new cases of invasive breast cancer and more than 48,000 non-invasive cases were diagnosed in the US. To put these figures in perspective, 64% of these cases are diagnosed early in the disease's cycle, giving patients a 99% chance of survival. Artificial intelligence and machine learning have been used effectively in detection and treatment of several dangerous diseases, helping in early diagnosis and treatment, and thus increasing the patient's chance of survival.*

Keywords: Breast Cancer, Machine Learning, Deep Learning, Classification, Bioinformatics.

الخلاصة : يعتبر السرطان من أخطر الأمراض على الإنسان، ولم يتم تطوير علاج دائم له حتى الآن. يعد سرطان الثدي أحد أكثر أنواع السرطان شيوعًا. وفقًا للمؤسسة الوطنية لسرطان الثدي، في عام 2020 وحده، تم تشخيص أكثر من 276000 حالة جديدة من سرطان الثدي الغزوي وأكثر من 48000 حالة غير غزوية في الولايات المتحدة. ولوضع هذه الأرقام في نصابها الصحيح، يتم تشخيص 64% من هذه الحالات في وقت مبكر من دورة المرض، مما يمنح المرضى فرصة للبقاء على قيد الحياة بنسبة 99%. وقد تم استخدام الذكاء الاصطناعي والتعلم الآلي بشكل فعال في الكشف عن العديد من الأمراض الخطيرة وعلاجها، مما يساعد في التشخيص والعلاج المبكر، وبالتالي زيادة فرصة بقاء المريض على قيد الحياة.

الكلمات المفتاحية : سرطان الثدي، التعلم الآلي، التعلم العميق، التصنيف، المعلوماتية الحيوية.

Introduction

Breast cancer is the most commonly diagnosed and leading cause of cancer deaths among women [1]. According to the World Health Organization (WHO), every year 2.1 million women have breast cancer worldwide. In 2018, an estimated 627,000 women died, representing about 15% of all cancer deaths among women. In the United States, it ranks first in the record of the most common cancers that women are expected to be diagnosed in 2019 at a rate of up to 30%. There are four types of breast tissue i.e., normal, benign, in-situ carcinoma, and invasive carcinoma. Benign tissue refers to a minor change in the structure of the breast, but it is not classified as cancer, and in most cases, it is not harmful to health. In-situ carcinoma remains in the mammary duct lobule system and does not affect other organs.[2]

Results and discussion

Many techniques or methods for cancer prediction and classification have been developed in the last years. In this section, comparing the results of implementing the proposed algorithm method with other related techniques. Which shows that the performance of the proposed method is more comprehensive than other methods.[3]

1.1. Te Data. We downloaded the METABRIC breast cancer dataset from the website Synapse (synapse.sagebase.org) and used the dataset's gene expression and clinical data. Te METABRIC dataset used in this study included 1980 samples. Moreover, according to the research work of Khademi, fve-year slot was used as the threshold to classify the two types of patients. Among them, 491 patients were divided into short-term survival samples and 1,489 patients into long-term survival samples. Meanwhile, the labels of short survival samples were set to 0 and the long life samples to 1. For gene expression data, handling methods of Sun et al. were used to preprocess it. Ten, the processed gene expression profle data and clinical data are normalized to between 0 and 1. For gene expression profle data, fve NMF algorithms are used to extract the features of the same dimension, especially with the feature dimension of 200. Each sample contained 25 dimensions of clinical information, such as age of diagnosis, tumor size, cancer grade, etc. [4]

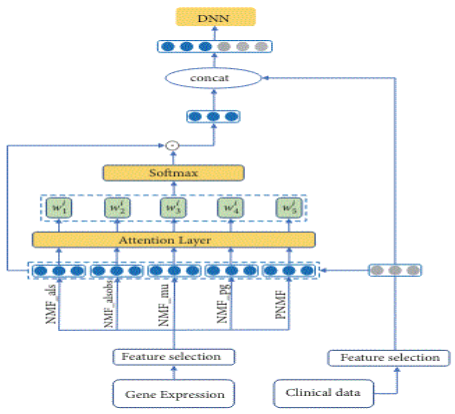


Figure 1: Te Overall Process of Our AMND Model for the Breast Cancer Prognosis Prediction [4].

Conclusion

Most papers published in the field of breast cancer detection and subtype classification use machine learning techniques. However, deep learning models have not been heavily investigated in this domain. This presents researchers with opportunities to use various deep learning mechanisms to predict patient status such as LSTM, GAN and RNN, as these types of research have not yet been conducted in the field.

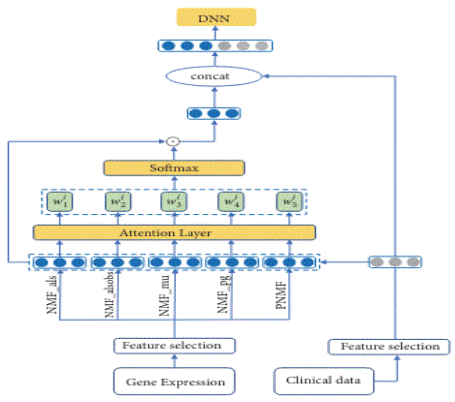
المقدمة

يعد سرطان الثدي السبب الأكثر شيوعًا للوفيات بالسرطان بين النساء [1]. وفقا لمنظمة الصحة العالمية، تصاب 2.1 مليون امرأة سنويا بسرطان الثدي في جميع أنحاء العالم. وفي عام 2018، توفي ما يقدر بنحو 627000 امرأة، وهو ما يمثل حوالي 15% من جميع وفيات السرطان بين النساء. في الولايات المتحدة، يحتل المرتبة الأولى في سجل أكثر أنواع السرطان شيوعًا التي من المتوقع تشخيصها لدى النساء في عام 2019 بنسبة تصل إلى 30 %هناك أربعة أنواع من أنسجة الثدي أي عادية، حميدة، في - السرطان الموضعي، والسرطان الغازي. تشير الأنسجة الحميدة إلى تغير بسيط في بنية الثدي، لكنها لا تصنف على أنها سرطان، وفي معظم الحالات لا تكون ضارة بالصحة. يبقى السرطان الموضعي في نظام فصيصات القناة الثديية ولا يؤثر على الأعضاء الأخرى.[2]

النتائج والمناقشة

تم تطوير العديد من التقنيات أو الأساليب للتنبؤ بالسرطان وتصنيفه في السنوات الأخيرة. في هذا القسم، مقارنة نتائج تنفيذ الطريقة الخوارزمية المقترحة مع التقنيات الأخرى ذات الصلة. مما يوضح أن أداء الطريقة المقترحة أكثر شمولا من الطرق الأخرى.[3]

1.1 بيانات تي. قمنا بتنزيل مجموعة بيانات سرطان الثدي METABRIC من موقع Synapse (synapse.sagebase.org) واستخدمنا التعبير الجيني والبيانات السريرية لمجموعة البيانات. تضمنت مجموعة بيانات Te METABRIC المستخدمة في هذه الدراسة 1980 عينة. علاوة على ذلك، وفقًا للعمل البحثي الذي أجراه الخادمي، تم استخدام فترة الخمس سنوات كحد أدنى لتصنيف نوعي المرضى. ومن بينهم، تم تقسيم 491 مريضًا إلى عينات البقاء على المدى القصير و1489 مريضًا إلى عينات البقاء على المدى الطويل. وفي الوقت نفسه، تم تعيين تسميات عينات البقاء القصيرة على 0 وعينات الحياة الطويلة على 1. بالنسبة لبيانات التعبير الجيني، تم التعامل مع طرق Sun et al. تم استخدامها للمعالجة المسبقة لها. عشرة، يتم تطبيع بيانات ملف تعريف التعبير الجيني والبيانات السريرية المعالجة إلى ما بين 0 و1. بالنسبة لبيانات ملف تعريف التعبير الجيني، تُستخدم خوارزميات fve NMF لاستخراج ميزات نفس البعد، خاصة مع بُعد الميزة البالغ 200. تحتوي كل عينة على 25 بُعدًا للمعلومات السريرية، مثل عمر التشخيص وحجم الورم ودرجة السرطان وما إلى ذلك. [4]



الشكل 1: العملية الشاملة لنموذج AMND الخاص بنا للتنبؤ بتشخيص سرطان الثدي [4].

الخاتمة

تستخدم معظم الأوراق المنشورة في مجال اكتشاف سرطان الثدي وتصنيف الأنواع الفرعية تقنيات التعلم الآلي. ومع ذلك، لم يتم التحقيق بشكل مكثف في نماذج التعلم العميق في هذا المجال. وهذا يمنح الباحثين فرصًا لاستخدام آليات التعلم العميق المختلفة للتنبؤ بحالة المريض مثل LSTM وGAN وRNN، حيث لم يتم إجراء هذه الأنواع من الأبحاث في هذا المجال بعد.

References المراجع والمصادر

1. Nassif, Ali Bou, et al. "Breast cancer detection using artificial intelligence techniques: A systematic literature review." Artificial Intelligence in Medicine 127 (2022): 102276.
2. Zhou X, Li C, Rahaman MM, Yao Y, Ai S, Sun C, et al. A Comprehensive Review for Breast Histopathology Image Analysis Using Classical and Deep Neural Networks. IEEE Access 2020;8:90931–56. https://doi.org/10.1109/ACCESS.2020.2993788.
3. Dhannoon BN. Predication and Classification of Cancer Using Sequence Alignment and Back Propagation Algorithms in Brca1 and Brca2 Genes. Int J Pharm Res 2019;11. https://doi.org/10.31838/ijpr/2019.11.01.062.
4. Chen H, Gao M, Zhang Y, Liang W, Zou X. Attention-Based Multi-NMF Deep Neural Network with Multimodality Data for Breast Cancer Prognosis Model. Biomed Res Int 2019;2019:9523719. https://doi.org/10.1155/2019/9523719.