Journal of Applied Artificial Intelligence
https://jaai.sabapub.com
ISSN: 2709-5908

2022 Volume 3, Issue 1 : 75 – 82
DOI : 10.48185/jaai.v3i1.468

Cancer Detection using Deep Learning: A Survey

Shuaib Ibrahim Odunlami[1,*], Edgar O. Osaghae[1], Basaky Frederick[1]

[1]Computer Science Department, Faculty of Science, Federal University Lokoja 058, Nigeria

Received: 03.04.2022 • Accepted: 21.05.2022 • Published: 30.06.2022 • Final Version: 30.06.2022

Abstract: Cancer refers to a group of diseases that are defined by abnormal cell proliferation and can spread to other parts of the body. Pancreatic cancer originates in the organ behind the lower stomach called the pancreas. The pancreas secretes digestive enzymes as well as hormones that help regulate sugar metabolism. Pancreas cancer is usually discovered late, spreads quickly, and has a poor prognosis. Cancer-specific symptoms may not appear until later stages, and there are no reliable screening techniques to identify high-risk patients. The manual method of checking whether a patient has Pancreatic cancer or not can be time-consuming and it can be tedious because it required experience in Radiologist interpretation; to avoid poor decisions by giving wrong diagnoses to patients. One of the leading causes of cancer death is pancreatic cancer; however, if found early enough, it can be treated. To solve this issue, an Artificial Neural Network (ANN) has been developed, trained, and tested using an abdominal Computed Tomography (CT) scan images dataset to detect pancreatic tumors by many authors. To detect the tumor, it employs image processing techniques of a convolution neural network to detect the tumorous area in the image after it has been pre-processed. It was observed that the technique gives better accuracy in tumor detection.

Keywords: Artificial intelligence, Neural Network, Computed Tomography, Machine Learning, Pancreatic cancer.

1. Introduction

The pancreas, an organ found behind the bottom region of the stomach, is where pancreatic cancer develops. The pancreas produces hormones that regulate blood sugar and appetite, raising stomach acids, and telling the stomach when to empty, as well as the digestive enzymes. Some tumors, both malignant and noncancerous, can affect the pancreas. Pancreatic ductal adenocarcinoma is the most frequent type of pancreatic cancer that starts in the cells lining the ducts that transport digestive enzymes out of the pancreas. Pancreatic cancer is rarely identified in its early stages when it is easiest to treat. This is because it, often, does not cause observable symptoms until after it has spread to other organs.

PC has a high death rate because cancer-specific symptoms only appear in more than 80% of patients at an advanced stage [1]. Pancreatic cancer (PC) is the world's fourth-largest cause of death in both men and women [2]. The PC-related death rate is believed to be rising, and by 2030, it is expected to overtake lung cancer as the second leading cause of cancer-related death. This is due to the lack of accurate screening technologies, the lack of sensitive and specific biomarkers, and the low prevalence of PC, identifying persons at high risk for early-stage disease is difficult [1].

* Shuaib Ibrahim Odunlami: ibrahim.shuaib@fulokoja.edu.ng

Several research has recently focused on the identification and validation of promising biomarkers for the early detection of cancer [2]. Furthermore, multiple clinical investigations have shown that Doppler Ultrasound (DU) can detect precancerous alterations in the pancreas in high-risk people. Endoscopic ultrasound (EUS), magnetic resonance imaging (MRI), computed tomography (CT) scan, or positron emission tomography (PET) scan are all options for diagnosing cancer [2]. Presently, diagnostic imaging techniques, such as CT, MRI, or EUS, can detect tumors as tiny as 0.5 cm. Some of these techniques are time-consuming and expensive, CT is the most efficient and accurate of the imaging techniques.

Early detection of pancreatic cancer is difficult due to the lack of an effective screening method to identify high-risk patients and the presence of cancer-specific symptoms occurs only at an advanced stage. To address this challenge, a deep learning model would be developed, using a CT scan dataset. The manual method of checking whether a patient has Pancreatic cancer or not can be tedious, time-consuming, and may even lead to the wrong diagnosis [3].

2. Literature Review

According to [4], Cancer is a term used to describe a group of diseases characterized by abnormal cell proliferation and the potential for spread to other parts of the body. Cancer, if left unchecked can grow out of control and develop tumors, which spreads into all other normal tissues that are close to it, through the bloodstream, and later affect other systems (digestive and respiratory) in the body. Thus, a malignant tumor can spread and destroy other tissues around it. Cancer can be broadly classified into four different categories. These include carcinoma (the kind that rises from the cell that helps to enclose and protect organs), sarcoma (type of malignant tumor of the bone), lymphoma (cancer that develops in the lymphatic system), leukemia (the cancer of the bone marrow and white blood cell) [5].

There are many cancers but some of the most dangerous ones are cervical cancer, breast cancer, lung cancer, skin cancer, and pancreatic cancer. Among cancer, symptoms are weight loss, new lump, bleeding, prolonged cough, a change in bowel movement, change of color in the skin, etc.

Machine learning (ML) which is a branch of artificial intelligence has many applications and can be used for cancer detection purposes. A neural network (NN) is a subset of machine learning that utilizes the CT scan for image processing to attain the desired result. NN is an information processing system designed to mimic the biological neural networking of the human brain. It consists of a processing element called neurons; each neuron connects to other neurons by a direct communication link which has its own weight [6]. Machine learning has therefore been utilized to aid in the diagnosis and detection of cancer [7].

For image processing, the neural network uses an architecture called Convolution neural network (CNN). CNN has three layers, the input layer which takes an input, the intermediate layer called the hidden layer, and the output layer where we get the trained output [8]. The close connection and the spatial formation between levels of convolution neural networks make them suitable for image processing and can make them automatically extract the rich correlative characteristic from the images [9].

Deep learning (DL) is also a subdiscipline of artificial intelligence that relies on artificial neural networks to extract patterns and make predictions from large data sets. The growing use of deep learning in healthcare, combined with the availability of well-characterized cancer datasets, has pushed research into deep learning's utility in understanding cancer's complex biology. While early

results are promising, the pancreatic cancer detection system is becoming more popular in the medical world with the move from manual detection of cancer to a more automated one with the use of machine learning.

Over the year a constant progression in cancer research has been carried out to find cancer types. Scientists used a variety of ways, and various symptoms to determine cancer types. With the introduction of new technologies in the medical field, a large amount of cancer data had been collected and is available to the medical research community. DL has helped in the speeding up of various significant scientific areas, including biomolecular structure prediction, gene discovery, genomics, and proteomics. Due to the efficiency and accuracy in the use of machine learning techniques in solving bioinformatics problems, its application in bioinformatics and related fields is widely accepted and continues to develop.

Cancer classification has long been a focus of cancer treatment planning and research. The traditional method of cancer classification is based mostly on the tumor's physical appearance. Various ways of detecting cancer in the body of a patient comprise different screening and tests to diagnose whether cancer is present. The different ways of detecting cancer include image testing, sputum cytology, and also by collecting tissue samples (biopsy). Computed Tomography (CT), Chest Radiography (X-ray), and Magnetic Resonance Images (MRI) scans are examples of imaging tests. With all the listed techniques cancer can be detected in a cancerous cell in the patient. However, there is the possibility that some enlarged cells, that is the non-cancerous tumor cell, may be confused as cancerous tumor cells (malignant).

The non-cancerous cell in the body of a patient is not harmful except when they are pressed against important organs in the body while the cancerous have to be detected early as to help the patient. The major drawback of the imaging test is that it requires a lot of manual analysis which leads to time-consuming and expensive costs thus making ML techniques a necessary tool for the early detection of cancerous cells.

Several investigations have demonstrated the potential inherent in the use of ML for predicting, identifying, and of course, detecting cancerous cells. For example, [10] a completely end-to-end deep learning (EE-DL) model for automatically diagnosing pancreatic cancers from original abdominal CT images has been developed. For locating pancreatic cancers from the original data, the methodology used in this model involves four steps: image screening, pancreas location, pancreas segmentation, and pancreatic tumor diagnosis. A dataset of clinical CT images from 319 patients was used to train the model, which was then tested on clinical CT images from 347 patients. The model provides fast and accurate pancreatic tumor diagnosis results, which may help surgeons make better surgical decisions.

[11] suggested that ANN has high discriminatory power for predicting pancreatic cancer risk and could provide an innovative technique for identifying people at advanced risk of pancreatic cancer who could benefit from more personalized screening and treatments. It's vital to have a better awareness of the disease's risk factors and symptoms so that healthcare providers and the general public can learn about possible preventative and/or early detection methods. It's critical to identify people at high risk who could benefit from screening for pre-malignant illnesses pancreatic intraepithelial neoplasia, intraductal papillary mucinous neoplasms, and mucinous cystic neoplasms are examples of mucinous cystic neoplasms, but there is currently no approved screening test.

[12] Random forests and support vector machine learning was used to predict high-risk intraductal papillary mucinous neoplasms (IPMN) of the pancreas using manually segmented CT

data. To predict IPMN risk, they evaluated pancreatic cyst and parenchyma regions derived from CT images in 103 patients. In order to create prediction models, clinical variables were integrated with imaging features. The best performance Area Under Curve (AUC = 0.81) was obtained when these traits were combined with the development of a preoperative model that can distinguish between low-risk and high-risk IPMN will help surgeons make better decisions.

[13] developed a method for screening pancreatic ductal adenocarcinoma (PDAC) by determining whether a sufficient number of voxels were segmented as tumors using a multi-scale segmentation algorithm. The framework has a sensitivity of 94.1 percent and a specificity of 98.5 percent, indicating that it has the ability to make a therapeutic Utility. However, the pancreas only takes up around 1.3 percent of each CT image in our CT dataset; the rest comes from other organs like the liver.

[14] presented a support vector machine system containing 24 guideline-based features and 385 radiomics high throughput features combined with regions of interest (ROI) marked by a radiologist to diagnose pancreatic serous cystic neoplasms (SCN). A CT scan of the patient was obtained, as well as information on pancreatic resection. A cross-validation cohort of 200 patients was used, whereas an independent validation cohort of 60 patients was used. Picture Archiving and Communication Systems were used to acquire demographic information, clinical information, and multidetector-row computed tomography images. An experienced radiologist manually retrieved the feature, which was then utilized to identify the most essential features using regression. The diagnostic model was created using a support vector machine classifier with 5-fold cross-validation. To validate the performance, an independent validation cohort was used. The proposed radiomics-based computer-aided diagnosis scheme could help clinicians make more accurate preoperative diagnoses and management decisions. However, it required an experienced radiologist for proper interpretation of the result.

[15] developed a U-net neural network architecture that can produce a better initial segmentation and is likely to be extended to work in a similar interactive deep learning segmentation approach. The U-net is fully trained to produce the best possible initial segmentation. The initial segmentation performance of U-net on CT dataset using dice similarity coefficient analysis. It was discovered that U-net provides a better baseline and can reach expert performance significantly faster than manual segmentation in the case of pancreas CT. This U-net architecture can provide a better solution for automatic medical imaging segmentation in general.

In another study, an artificial neural network (ANN) was developed, trained, and tested using data from the National Health Interview Survey (NHIS) which consists of pancreatic, lung, colorectal, and ovarian cancer datasets, which together contain 898 patients diagnosed with pancreatic cancer. ANN was used to analyze individual pancreatic cancer risk prediction. This is because ANN has high discriminatory power that it could be used as a novel tool for identifying persons at higher risk of pancreatic cancer who would benefit from more individualized screening and therapy [16].

[17] identified malignancies by employing CT imaging. To find the tumor image processing techniques and a rudimentary classifier was used. A minimal distance classifier was employed to detect the tumorous area after the image had been pre-processed. The detection accuracy of cancers is roughly 60%. The goal of this study was to improve the prognosis by establishing the best appropriate early detection procedures for this aggressive cancer. CA 19-9 is the most broadly utilized marker at the time, and while it appears to be appropriate for both diagnosis and follow-up, its sensitivity and specificity remain poor [18].

[19] developed a technique that makes use of a convolutional neural network to detect whether the input lung CT picture contains cancer tissue. CT scans of lung cancer in various shapes and sizes were used to train the system. The system was able to detect the presence or absence of cancer with a 96 percent accuracy. The dataset utilized in the study was classified with a 93 percent accuracy using a multilayer perception network backpropagation method employing the GLCM feature.

[20] developed an approach for detecting lung cancer using image segmentation. Region growth, marker-controlled, watershed, and marker-controlled with masking are some of the image segmentation techniques employed. Following the detecting phases, picture augmentation with the Gabor filter and feature extraction is performed. The watershed with the masking method, which has great accuracy, was found to be the optimum option.

[21] emphasize deep learning applications for genomics data types, such as genomic, methylation, and transcriptome data, as well as histopathology-based genomic inference, and offer ideas on how the various data types might be combined to construct decision support tools. They show how deep learning can be used in cancer diagnosis, prognosis, and treatment management using particular examples. However, there were certain drawbacks, such as a shortage of rich data and the need for more understandable deep learning models.

The foregoing reveals the different works, which are by no means exhaustive, that employ machine learning algorithms and modern scientific approaches to tackle challenges linked to medical diagnostic issues. A consensus is that, for pancreas cancer, early detection of malignant tumors necessitates image-based databases which can be time-consuming and error-prone if done manually. The finest deep learning technique for early diagnosis of pancreatic cancer is neural networks, especially when used with picture datasets for categorization. Convolutional Neural Networks is the most commonly used forms of Neural network in image categorization since they allow for a variety of designs to be constructed on top of them.

In the following section, a possible procedure for image detection using CNN is discussed for possible application.

3. Methodology

This section describes the methods used to achieve the objectives of a proposed system, the development of which consists of the following five stages.

- Business understanding- here the focus is on the project objective of early detection of pancreatic cancer.
- Data understanding- this phase deal with the curated dataset for the pancreas by trying to understand it and detect interesting fact about the dataset.
- Data preparation- this phase deal with creating a final dataset that contains annotation of Images. The annotation gives details about the area of interest in the image, which is a cancerous tumor.
- Modeling involves the development of the cancer detection model by using the selected technique to train the dataset.
- Evaluation- in this phase, the accuracy of the model is tested with the dataset.

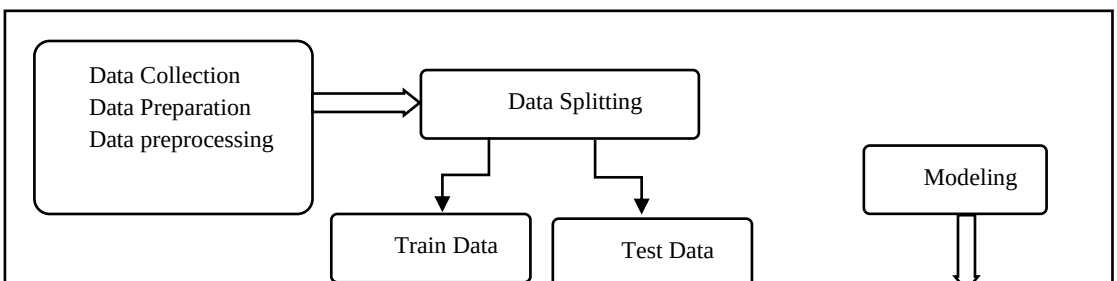

Figure 1. shows the iterative process of the methodology

4. Proposed System

The models of the proposed system were designed, including the process models and data models. The model will be built using a Convolutional Neural Network (CNN) technique, which is a type of Artificial Neural Network and a major player in the field of computer vision. The images will contain slices of pancreatic cancer from the CT scan and the annotations will help identify the area which contains the tumor in the scanned images. The goal of a convolutional layer is to filter through the image that passes through it, as it moves over the image to check if a pattern is present. The proposed model will be able to detect pancreatic cancer, a procedure that is been done manually in the past, and hence it will be of immense help for medical practitioners. The model has many advantages like detecting smaller nodules in scan images which will help save patient lives before it grows to uncontrollable size, thus reducing the death rate caused by pancreatic cancer.

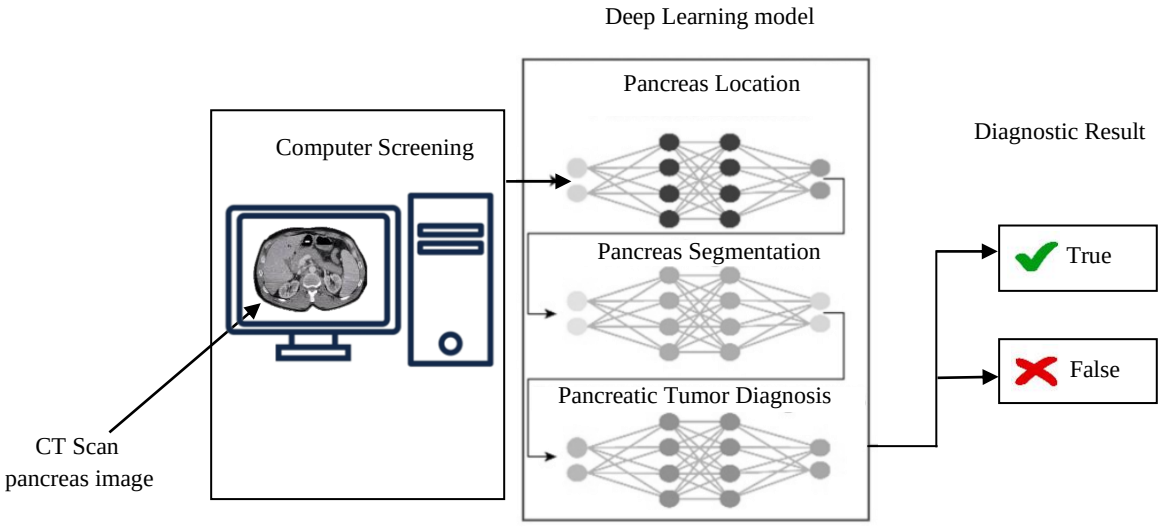

Figure 2. Proposed System

Figure 2 describes the activity and depicts the functionality of the system. It shows the design of a pancreatic cancer detection model which consists of the detective analysis module and cancerous tumor detection.

5. Conclusion

Pancreatic cancer is one of the most lethal cancers. Due to the lack of distinctive early cancer symptoms, it is frequently detected at an advanced stage in recent years. Because there is no sure cure for cancer in its advanced stages, early detection through regular diagnosis is critical in the fight against cancer. Deep learning in this aspect, appears to be a viable option. Deep learning-based technologies are less expensive and faster than most other methods for detecting malignant tumors and establishing their type. The proposed method, on the other hand, is expected to provide an efficient and accurate preoperative diagnosis to aid patient clinical management.

References

[1] S. Boursi, B. Ben, F. Brian, G. Bruce, H. Kevin, et al. “Clinical prediction model to assess risk for pancreatic cancer among patients with pre-diabetes”. Oncol. suppl.e1 March, 2017.

[2] P. Klein, S. Lindström, B. Mendelsohn, E. Steplowski, A. Arslan, et al. “An absolute risk model to identify individuals at elevated risk for pancreatic cancer in the general population". PLoS ONE. 8(9), September, 2013.

[3] V. Dasale, D. Vijaykumar, T. Bakhale, S. Kadam, D. Somnath “Detection of lung Tumor in its early stages using Image processing Techniques”. (IJCST), 5(2) March, 2017.

[4] Cooper GM. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. The Development and Causes of Cancer. Available from: ttps://www.ncbi.nlm.nih.gov/books/NBK9963

[5] Y. Nemlich, E. Greenberg, R. Ortenberg, M. Besser, L. Rivkin, et al. “MicroRNA-mediated loss of ADAR1 in metastatic melanoma promotes tumor growth”: *J Clin Invest* 123(6), June 2013

[6] M. Hussain, M. Tabassum, P. Ansari, S. Gawas, N. Chowdhury “Lung cancer Detection Using Artificial Neural Network and Fuzzy Clustering.” IJARCCE, 4(3) March, 2015

[7] J. McCarthy, K. Max, P. Hoffman, G. Alexander, O. Philip “Application of machine learning and high dimension visualization in cancer detection, diagnosis, and management.” Annals of New York Academy of Sciences, 239-262 May 2004

[8] D. Jaswal, V. Sowmya, K. Soman “Image Classification Using Convolutional Neural Networks”: IJART, Volume 3, Issue 6, June-2014

[9] X. Han, & Y. Li “The Application of Convolutional Network in handwritten Numeral Recognition.” IJD: Theory and application.4(3). 2015

[10] K. Si, Y. Xue, X. Yu, X. Zhu, Q. Li, et al. “Fully end-to-end deep-learning-based diagnosis of pancreatic tumors.” Theranostics, 11(4) 2021

[11] M. Andrew, K. Paul, C. Richard, J. Claire, G. Helen, et al. “Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment, and outcomes.” *Baishideng Publishing Group Inc.* 24(43):4846-4861 November 2018.

[12] J. Chakraborty, M. Abhishek, G. Lior, A. Marc, L. Liana, et al. “CT radiomics to predict high-risk intraductal papillary mucinous neoplasms of the pancreas.” Med Phys. 45(11): 5019-29 November 2018.

[13] Z. Zhu, Y. Xia, L. Xie, E. Fishman, L. Yuille, “Multi-scale coarse-to-fine segmentation for screening pancreatic ductal adenocarcinoma.” MICCAI, 2019; 11769: 3-12.

[14] R. Wei, K. Lin, W. Yan, Y. Guo, Y. Wang, et al. “Computer-aided diagnosis of pancreas serous cystic neoplasms: a radiomics method on preoperative MDCT images.” Technol Cancer Res Treat. January 2019, 1(18).

[15] T. Boers, Y. Hu, E. Gibson, D. Barratt, E. Bonmati, et al. “Interactive 3D U-net for the segmentation of the pancreas in computed tomography scans”, Phys Med Biol. 65(6) March 2020.

[16] M. Wazir, H. Gregory, N. Bradley, F. James, J. Kimberly et al. “Pancreatic Cancer Prediction through an Artificial Neural Network.” Department of Therapeutic Radiology, Department of Internal Medicine, School of Medicine, Yale University, New Haven, CT, United States. Artif. Intell., May 2019 | https://doi.org/10.3389/frai.2019.00002.

[17] S. Jeenal, S. Surve, V. Turkar. “Pancreatic Tumor Detection Using Image Processing.” Fr. Conceicao Rodrigues College of Engineering, Bandra(W), Institute of Technology, Wadala€, Mumbai-400037, India 2015.

[18] N. Bacalbasa, A. Gireada, I. Balescu. "Tumor markers in pancreatic cancer", Scholarly Journal Vol. 7, Iss. 2: 75-78 Jun 2015.
[19] S. Sasikala, M. Bharathi, R. Sowmiya "Lung cancer detection and classification Using Deep CNN", IJITEE, 8(25)-262 December, 2018.
[20] A. Mazin, A. Mohd, I. Raed, K. Mohamad, A. Dheyaa "Automatic segmentation and automatic seed point selection of nasopharyngeal carcinoma from microscopy images using region growing based approach", JCSV vol 20, Pg 61-69 May 2017.
[21] T. Khoa, K. Olga, B. Andrew, W. Elizabeth, P. John et al. "Deep learning in cancer diagnosis, prognosis, and treatment selection." Genome Medicine 13(1) :152. September, 2021.