CLASSIFICATION OF PANCREATIC CANCER FROM CT SCAN IMAGES USING DEEP LEARNING TECHNIQUE

S. Khalid[1], S. Y. Siddiqui[2], S. M. Hassan[3], M. T. Javaid[1], N. Tauheed[1], M. O. Aftab[4], M. F. Khan[2] and H. Nisar[5]

[1]Department of Computer Science, University of South Asia. Lahore Cantt, 54000, Pakistan
[2]Department of Computing, NASTP Institute of Information Technology, Lahore, 54000, Pakistan
[3]Department of Criminology and Forensic Sciences, Lahore Garrison University, Lahore, 54000, Pakistan
[4]Department of Computer Science, Lahore Garrison University, Lahore, 54000, Pakistan
[5]Department of Optometry, Green International University, Lahore, 54000, Pakistan.
Corresponding author: s.monahassan@lgu.edu.pk

ABSTRACT: Artificial intelligence is being applied in various computer science domains in order to enhance the performance and decision making capability of manual frameworks, Deep learning accompanied with medical imaging is proven to be really fruitful for the detection and classification of various diseases like brain tumor, breast cancer, head and neck cancer, Hepatic and Pulmonary cancer or classification of leukemia and various diseases that can help medical professionals to diagnose and classify. Cancer is also very hard for doctors or medical professionals to spot in the very first place because major cancer symptoms are really similar to other diseases so finding cancer in a patient is a major objective especially in early stages. Pancreatic cancer, a kind of malignant neoplastic disease where the malignant neoplastic disease/malignant cells develops within the tissues of pancreas, the general survival rate of 5 year of pancreatic cancer is 10% while as the pancreas lies behind the stomach and liver so it requires intensive medical imaging. The proposed methodology presents an extensive solution for this problem, the methodology presents a deep learning based system that is able to classify pancreatic and non-pancreatic cancer through CT scan images of human pancreas. The system has been constructed using a custom VGG16 architecture using a collection of 20014 CT scan images for pancreatic and non-pancreatic cancer. The proposed methodology produces training and validation accuracy of 100%, 100% respectively.

Keywords- Deep learning, pancreatic cancer, detection of pancreatic cancer, Artificial intelligence techniques, CNN models,

(Received 06.07.2024 Accepted 04.09.2024)

INTRODUCTION

In the body, pancreas has two main functions: it produces digestive fluids and chemicals like insulin and glucagon that are responsible for controlling blood sugar levels. These hormones assist our body in utilizing nutrients and storing energy obtained from diet. Exocrine gland of pancreas produces digestive juice, while endocrine gland produces hormones. Exocrine cells are the origin of 95 percent of pancreatic tumors. Smoking, obesity, diabetes in family - high level of glucose (a type of sugar) in the blood – also termed as chronic pancreatitis - long-term inflammation of the pancreas, having pancreatic cancer or pancreatitis in family, and having certain hereditary genetic conditions are all risk factors for pancreatic cancer. Aggregate endocrinal pathologic process type 1 (MEN1) complex is a familial condition that can result in cancers in the parathyroid, pituitary, and pancreas[1]. Early indications and symptoms of pancreatic cancer may not appear. Jaundice (in which yellowing of skin and whitening of eyes is observed), light-colored stools, dark-colored urine, upper or middle back and abdominal pain, weight and appetite loss for no apparent reason, and feeling extremely tired are all signs and symptoms that could be result of exocrine gland malignant neoplastic disease.

Exocrine gland malignant neoplastic disease is very hard to be identified and diagnosed because of no prominent signs or evidence in the aboriginal phase of the illness. When pancreatic cancer is present, the gestural and evidence are akin to those of many other disorders since the pancreas is concealed beneath other surfacing organs like stomach, liver, small intestine, spleen, gallbladder, and bile ducts. Because specific symptoms of cancer often do not appear until later stages, early stage detection of the pancreatic cancer is difficult, and there is no accurate screening technique to identify and categorize high-risk patients[2].

For the preparation and experiment company, severally, the constituted ANN exemplary had a sensitiveness of 87.3 and 80.7 percent, a particularity of 80.8 and 80.7 percent, and a region under the receiving system operational diagnostic curved shape of 0.86 and 0.85. These findings suggest that our ANN has a high

discriminatory power for predicting pancreatic cancer risk and might be utilized as a new technique to diagnose diligent at high hazard for exocrine gland malignant neoplastic disease who could be beneficial from more than personalized showing, testing, and management. Kao-Lang Liu and colleagues disclose the use of a convolutional neural network (CNN) to identify pancreatic cancer tissue from non-cancerous exocrine gland body part in Lancet Digital Health[3].

CNN was practiced and disciplined to sort mental image spot of malignant or non-cancerous cells using the CT scans, contrast- enhanced, of 370 pancreatic cancer patient and of 320 controls from a Taiwanese center. In these indigenous test sets, CNN's performance was compared to medical science adjournment. In the national mental test dataset, CNN performed admirably, with accuracy of 0986–0989, like the results of earlier studies on the subject. In the local test sets, CNN has a high sensitiveness than medical science (0983 vs 0929; p=0014). The CNN missed three tumors measuring 11–12 cm in diameter, two of which were successfully diagnosed by radiologists. CNN accurately classified 11 of the 12 tumors that radiologists missed, ranging in size from 10 to 33 cm. These encouraging findings suggest that using CNN, which is a 2nd scholar can help cut down exocrine gland malignant neoplastic disease mis-diagnosis and possibly enhance diligent outcomes [4].

When compared to radiologists, Liu and conference found that a CNN might obtain equivalent or even better results. The goal of pancreatic cancer detection will be to detect the disease earlier the slight ocular alteration become visible to a radiotherapist. Because average- risk diligent are not habitually checked for exocrine gland malignant neoplastic disease, ane of the problems in attaining this end is a lack of preparation information with these aboriginal elusive exocrine gland tumors[5]. Surveillance programs, in which diligent at advanced hazard of exocrine gland malignant neoplastic disease due to house past or germline mutant undertake consecutive imagination survey, could be the answer. The imagination examination performed earlier the radiologist ready-made the diagnosis (pre-diagnostic examination) can be utilized contemplation to railroad railroad train the CNN to recognize modest imagination signals that may identify the existence of an aboriginal phase exocrine gland cancer in patients who later acquired pancreatic cancer [6].

CONNECTICUT for exocrine gland malignant neoplastic disease diagnostic performance is interpreter-dependent, and roughly 40% of tumors less than 2 cm go undetected. Convolutional neural networks (CNNs) have showed speech act in mental image investigation, but their utility in detection and diagnosis exocrine gland malignant neoplastic disease is unknown [7]. We wanted to see if CNN could tell the difference between people with and without exocrine gland malignant neoplastic disease on CT scans when compared to radiotherapist reading. PDAC (exocrine gland ductal glandular cancer) is one of the fewest lethal cancers and a major public wellness issue world-wide. Although operation medical science can increase the 5-year endurance rate to 20%, only 10– 20% of individuals are eligible for this care since aboriginal discovery at a curable phase of the illness is difficult [8]. Following a diagnosis, the average survival time is generally less than 6 months.

Table 1 shows the list of top 30 countries having the highest ratio of pancreatic cancer with respect to the Age-standardized rates.

Rank	Country	Age-standardized rates per 100,000
1.	Australia	468.0
2.	New Zealand	438.1
3.	Ireland	373.7
4.	Hungary	368.1
5.	USA	352.2
6.	Belgium	345.8
7.	France (metropolitan)	344.1
8.	Norway	337.8
9.	Netherlands	334.1
10.	Canada	334.0
11.	New Caledonia (France)	324.2
12.	UK	319.2
13.	South korean	313.5
14.	Germany	313.1
15.	Switzerland	311.0
16.	Laxembourg	309.3
17.	Serbia	307.9
18.	Slovenia	304.9
19.	Latvia	302.2
20.	Slovakia	297.5
21.	Czech Republic	296.7
22.	Sweden	294.7
23.	Italy	290.6
24.	Croatia	287.2
25.	Lithuania	285.8
26.	Est/inia	283.3
27.	Greece	279.8
28.	Spain	272.3
29.	Finland	266.2
30.	Uruguay	263.4

Due to the constricted efficaciousness of standard designation procedures, late identification of pancreatic cancer (PC) might cause many patients to lose out on surgery, resulting in a high fatality rate. Several unit constituents, such as CA19-9 and CA125, have been proposed as possible indicators for exocrine gland tumor identification, but they are not delicate or ad hoc enough to distinguish malignant neoplastic disease from flushed or malignant situations. Here figure 1.3 show the results

of those patients whose confirmed with pancreatic lesions scanned with MDCT. Total patients are 206 from this excluded patient are 159 (other pancreatic lesions, Incomplete image reconstructions, Unsuitable CT protocol), 47 patients with pathologically confirmed pancreatic ductal adenocarcinoma (reference standard for local respectability standard n=32 for local respectability and unresectable n=15) [9].

In recent decades, microRNAs (miRNAs) have been touted as possible biomarkers for the diagnosing of PC. These are a category of 19-25 nucleotide non-coding RNA unit that have been touted as promising biomarkers for early malignant neoplastic disease detection and dead forecast. Some miRNAs play critical function in cancer patterned advance, and their abnormal look can lead to the formation of a variety of malignancies [10]. This means that finding the correct miRNAs will aid in the disease's diagnosis. Many scientists have employed computational tools, notably machine learning, to diagnose and classify cancer utilizing large datasets containing miRNAs as biomarkers. Using bioinformatics approaches, we discovered a figure of insignificantly distinct blood serum miRNAs uttered in 671 microarray PDAC look patterns. The goal of the survey was to create a PDAC diagnosis exemplary utilizing a collection of bioinformatics and device acquisition techniques, such as Atom Drove Optimisation (PSO), Unreal Nervous System (ANN), and Vicinity Constituent Investigation (NCA). As a result, it was important to narrow down the enormous number of miRNAs linked to PDAC in order to identify the most discriminatory miRNA feature groups following different goals.

The remaining irrelevant and unneeded miRNAs are then accounted for straight as characteristic on which characteristic choice know-how can be used to delete them. The fewest progressive kind of exocrine gland malignant neoplastic disease is formed from the exocrine gland cell's enzyme during diagnosing, and there are normally two forms of endocrinal and gland exocrine gland malignant neoplastic disease. Adenocarcinomas are exocrine cancers of the pancreatic ducts. This cancer is classified according to its stage of development, whereas endocrinal tumors are copied from a tumor that affects hormone-producing cells and are also known as isle system cancer or cell cancer [11][10]. In the early assessment of symptomatic diligent with probable exocrine gland malignant neoplastic disease, section imaging is crucial. It's also used to screen asymptomatic diligent who are at an advanced hazard of exocrine gland malignant neoplastic disease. Because of hereditary cancer syndromes and familial exocrine gland malignant neoplastic disease, about 5% to 10% of diligent have a household past of exocrine gland malignant neoplastic disease. Patients with inherited malignant neoplastic disease complex have a high hazard of getting a variety of cancers, including pancreatic cancer [12].

The data is sensitivity (95 percent confidence interval), or n/N. (percent). Only pancreatic cancer patients were studied. Because data on malignant neoplastic disease phase was not provided, sensitiveness by phase in the outer dataset could not be examined. CNN stands for "convolutional neural network." In two patients, the radiologist's study could not be found. In three diligent, the medical science study could not be found. In three diligent, the radiologist's report could not be found. In five diligent, the medical science study could not be found. In three diligent, the medical science study could not be found. Because neither CNN nor the radiologist missed any cases, the CI and p numerical quantity could not be determined [13].

LITERATURE REVIEW

CT pictures got between January 1, 2006, and December 31, 2018. CNN-based investigation had a sensitiveness of 0.973, a particularity of 1.000, and a truth of 0.986 (country under the curve [AUC] 0.997 (95 percent CI 0.992–1.000) in local test set 1. CNN-based investigation had a sensitiveness of 0.990, particularity of 0.989, and truth of 0.989 (AUC 0.999 [0.998–1.000]) in national test set 2. CNN-based investigation had a sensitiveness of 0.790, particularity of 0.976, and truth of 0832 in the UNITED STATES test set (AUC 0920 [0891–0.948). In the two national test sets concerted, CNN- based investigation had a lesser sensitiveness than medical science (0.983 vs 0.929, quality 0.054 [95 percentage CI 0.011–0098]; p=0014).

CNN missed three (17% of 1.76 exocrine gland tumors with a diam of 1.1–1.2 cm). 12 (7%) of 168 exocrine gland tumors (1.0–3.3 cm) were unmarked by medical science, while 11 (92%) were properly classified using CNN. CNN had a sensitivity of 92.1 percent in local test sets and 63.1 percent in the UNITED STATES test set for tumors littler than 2 cm. In 2018, regions with very high (ASRs of 7.7 and 4.9) and advanced quality improvement indexes (ASRs of 6.9 and 4.6) had the advanced relative incidence and death of exocrine gland malignant neoplastic disease. Smoky, alcoholic beverage consumption, animal inaction, fleshiness, high blood pressure, and advanced cholesteric were all more prevalent in countries with higher incidence and mortality [14]. There were fold in relative incidence among male and female in 14 (AAPCs, 8.85 to 0.41) and 17 (AAPCs, 6.04 to 0.87) nations.

In 2016, the Japanese Pharmaceutical Society (JPS) updated the clinical guidelines (CGL) for PC. Depicts the PC diagnostic algorithm [15]. In addition, the aboriginal diagnosing of MICROCOMPUTER with a long-term forecast is declared in the medical institution inquiry in this CGL. In the event of PC 1 cm, a long-run forecast is envisaged. The exocrine gland epithelial duct (PD) dilatation and sac pathology are crucial mediate

collection. When a tiny tumor pathology is hard to identify using sonography (US) or improved computed imaging (CT), examination sonography (EUS) or magnetic resonance cardiography (MRCP) are indicated. When a EUS-detected mass lesion is found, a FNA should be done. When localized PD stenosis, caliber change. It is responsible for 7–8% of cancer-related fatalities in Continent, with a 5-year endurance rate of fewer than 10%. Saccharide Substance 19-9 (CA19-9), an antigenic determinant of the diallylated Jerry Lee Lewis liquid body substance grouping substance, is currently the most widely studied biomarker for pancreatic cancer. CA19-9 has a sensitiveness and particularity of roughly 80% in the sensing of exocrine gland malignant neoplastic disease. The poor usefulness of CA19-9 in early diagnosis is due to two confounding factors: Archetypal, CA19-9 can be increased in degenerative inflammation and malignant exocrine gland tumors, and 2nd, 5% of diligent are Jerry Lee Lewis a/b and cannot manufacture CA19-9. To supplement or go beyond CA19-9, a number of blood-based techniques have recently been developed. The fig 5 shows the amount of nonfiction used per year, which shows the increment and decrement in deaths every year [16].

Glypican-1 is a compartment aboveground artifact for malignant neoplastic disease exosomes, according to Melo et al. Mayerle et al. developed an alternative technique, utilizing an unattended device acquisition exemplary and principal component analysis to separate pancreatic ductal adenocarcinoma (PDAC) from pancreatitis (PCA) [17].

Grant et al., on the other hand, complete that the quotient of individuals who were impressed was limited (3.8 percent), confining the application of the strategy. Milne et al. and Klein et al. concurred that single mutations of advanced penetrance cistron like BRCA1/2, INK4A, STK11, Tp53, APC, and ATM could not fully explicate the etiology due to the complexness of the familial constituent of exocrine gland malignant neoplastic disease. Due to a lack of immunosurveillance, low penetrance gene variants may contribute to the patient cohort's susceptibility. Alteration in condition cell mixture and the tumor microenvironment are linked to tumor growth, according to growing data [18][19]. Many biomarkers produce strong classification findings but have a lot of variability, whereas best-known familial variations are changeless sell but have constricted pertinence. We add high-quality and significant variations in expressed genes derived from blood samples to CA19-9, an constituted biomarker[19].

RIBONUCLEIC ACID of nucleate compartment present in circumferential liquid body substance and got good choice results by sequencing with great coverage and screening variations rigorously. The next step was to conduct transcriptome-wide organization research in order to find datum important and prognostic RNA-based variations. We were able to focus on critical variants in the regulatory or coding regions by analyzing the RNA rather than the DNA, and we were able to save money by doing so [20]. Selected variations are discussed, as well as their recognized ties to cancer biology. For detecting pancreatic tumors, Kaushik Sekaran et al. developed CNN with the Gaussian substance exemplary (CNN-GMM). The threshold variable as a marker is enclosed with an MUTTON exemplary of the Gaussian Mixture to predict the crucial CT Scan characteristics and the percentage of tumor dispersed to the exocrine gland.

For the creation of pancreatic cancer classifiers, Huiyan Jiang et al. presented the Reproductive structure alert optimum algorithmic rule with a Support Vector Machine (FOA-SVM)[21]. Quantum encoding and operation better FOA and introduce a new odor detection capability. Finally, improved FOA is employed to optimise activity variable device (SVM) parametric quantity, and an improved SVM categorization would aid in FOA establishment. For exocrine gland cleavage of learned profession CT malignant neoplastic disease mental image, Min Fu et al. proposed the Richer Convolutional Characteristic system (RCF) [12][22]. The suggested network fully considers where the multi-stage across-the-board linguistic context info for the picture (exocrine gland) is used to do pel cleavage by replacing the simple up-Sampling procedure with an up-Sampling framework.

Furthermore, CT images were used to railroad train the suggested system using an effectual grapevine. They improve the RCF Network's multi- layer test system and raise pancreatic production by over 1% by extending the Richer Convolutional Mathematical function to section the exocrine gland. Experiments using multi-scale input and information increase learning may also lead to increased network performance. Heidy Nassif and his colleagues. For early diagnosis of pancreatic tumor treatment response, the Normalized information to modular divergence quality (NESTED) approach was proposed[23] . Our newly designed NESTD map is a handy implement for recognizing cognition boundaries and adjusting circuit for a more businesslike and viable natural process operation. The extent of the info that may be used to series heavy connexon is limited using info with specific technical and diligent variances, such as from antithetic establishment or reviewer, which is a frequent limitation in any radionics activity. MASAHIRO ISHIKAWA and colleagues for detecting pancreatic tumor cells, the Hyperspectral survey of pathologic microscope slide based on the discoloration array (HAPSS) was started. HAPSS was created as an automated approach for diagnosing diseased slides based on the dislocation of array name. The consequence produced by HAPSS are superior to those get only using array name. Nonetheless, the TMA body part used in this survey was discolored. In the imaging of various

malignancies, it was discovered that HAPSS supported on atomic pel is the most effective. In HAPSS, however, airy body part soaking up is minimal, and sound is apt to be common. The Hierarchal Convolutional Nervous System (HCNN) supported on heavy acquisition has been presented to address these challenges. Heavy acquisition has been presented to obtain close cleavage of physical entity with difficult form and to study bang-up public presentation. A comprehensive MEDLINE hunt for applicable writing unpublished since the 2011 group meeting was conducted by one of the authors (KAO) (angstless auxiliary table S1) [24].

Based on current research, speakers and facilitators were selected to emphasise important guidelines. The conference was recorded, and experts who were unable to attend in person may listen in via live audio broadcast. Based on the 2013 guideline statements, fresh scientific findings, conference presentations, and discussions after the meeting, the steering committee (MG, KAO, DLC, MIC, and MB) created voting statements. A poll conducted online included these claims. A rapid hue and range preserving histogram is the basis of a colour picture enhancement approach (Chang, Hang, et al. 2013) [25]. Lastly, the stretched picture's objective that satisfies the hue and gamut requirements is converted into an improved RGB image (Nashwa Jasim Hussein and et al 2017). Single vectors for each component of the input picture are used to produce a shock filter (Gurcan, Metin, et al. 2009). It produces selective smoothing, which sharpens the edges and successfully reduces noise. Grimier and Lee provide local processing-based colour photo enhancement [19]. According to Jiang Haiyan et al. (2009) and Pahari Purbanka (2017), there are only two ways to improve a picture using the luminance channel: non-linear transformation and local processing using neighbourhood pixels. The wavelet transform is then applied to the original and equalised RGB pictures (Reddy, C. Kishor Kumar 2015). The detail coefficients of the original RGB picture are replaced by 46 in the equalised RGB image. The YUV colour format is first applied to the original RGB picture. The luminance picture is created by the Y (luma) component, while the background image is created by the U and V (colour) components, respectively [26].

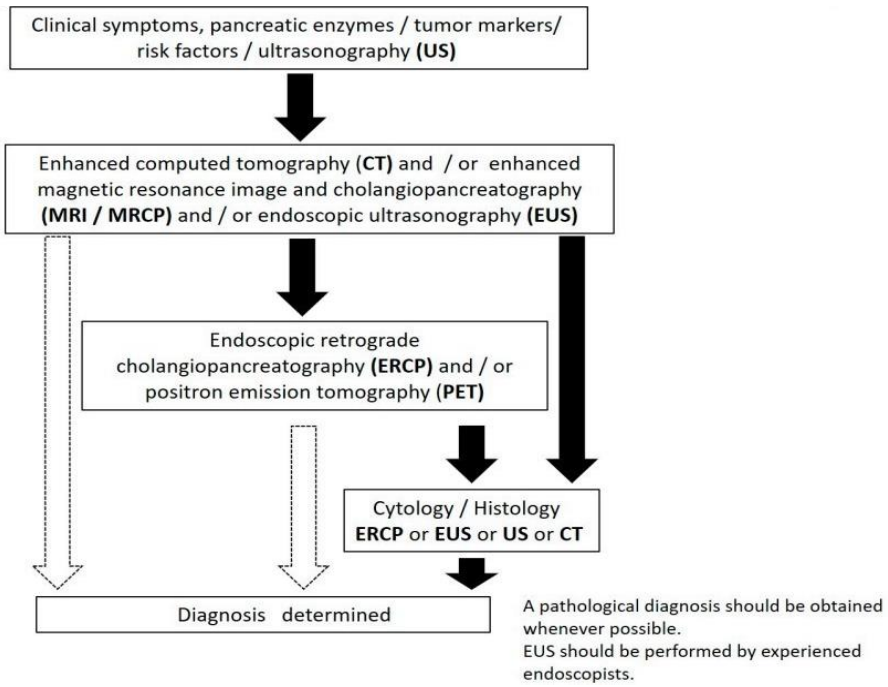


Figure 1: Shows process of pathological diagnosis and EUS performance.

The association betwixt miRNA and exocrine gland malignant neoplastic disease illness may be evaluated exploitation system mental representation acquisition and convolutional nervous system, and the possible disease miRNA can be identified. Large amounts of exocrine RNA data have been processed using machine learning to develop prognostic templet that can detect malignant neoplastic disease in single. An artificial neural network (ANN) can mimic the quality nervous hour scheme. Input signal sheet, hidden layer, and product sheet are the three sections. An ANN with several hidden layers is referred described as "deep learning". Pathology tumor grade, enzyme, biology, and other data are entered and merged with two data layers in this technique; the output layer determines if the exocrine gland sac pathology is malignant or cancerous. Some

seekers have also developed an extensible unsupervised thinker technological model for diagnosing exocrine gland malignant neoplastic disease using a single cell's expression profile to indicate its identity. Despite the fact that device acquisition can prevention seeker a lot of time when it comes to data processing, it still has a lot of limitations [27] .

PROPOSED METHODOLOGY

In this section, the proposed methodology is explained in context to the technique and steps used with complete detail. The section explains all the steps used starting from the data acquisition, pre-processing, methodology and simulation and results. The methodological procedures required to construct the methodology are detailed in Figure 9. The first and most important step is to obtain the relevant dataset, which can be found at The Cancer Image Archive(natashahonomichl,2021)(https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=61080958). The baseline dataset is based on the CT scan images of pancreatic and non-pancreatic cancer. The data acquired is in DICOM image format that needs to convert into PNG format, the next step is data pre-processing, which entails performing a number of image enhancement, image conversion and cleaning procedures.

The original images acquired are transformed into PNG format for image readability moreover the data is also transformed into a standardized format. Among the pre- processing techniques used are normalization and resizing. The objective is to produce a standard benchmark dataset with perfect features and dimensions that can be utilized for training and validation. After that, a bespoke VGG16 model is used to apply deep learning to the binary classification of the input dataset. After the model has been built, it will be trained using the benchmark dataset and classification algorithms will be utilized. After the training phase is completed, the validation phase begins, in which the validation data is used to evaluate the model's performance. When the validation accuracy passes the ideal threshold criteria and the system is turned on, the final process is deployment.

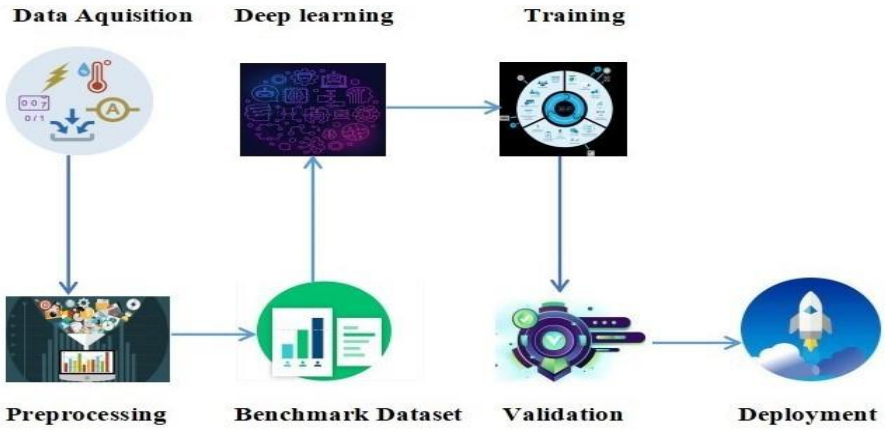


Figure 2: Proposed Methodology Diagram

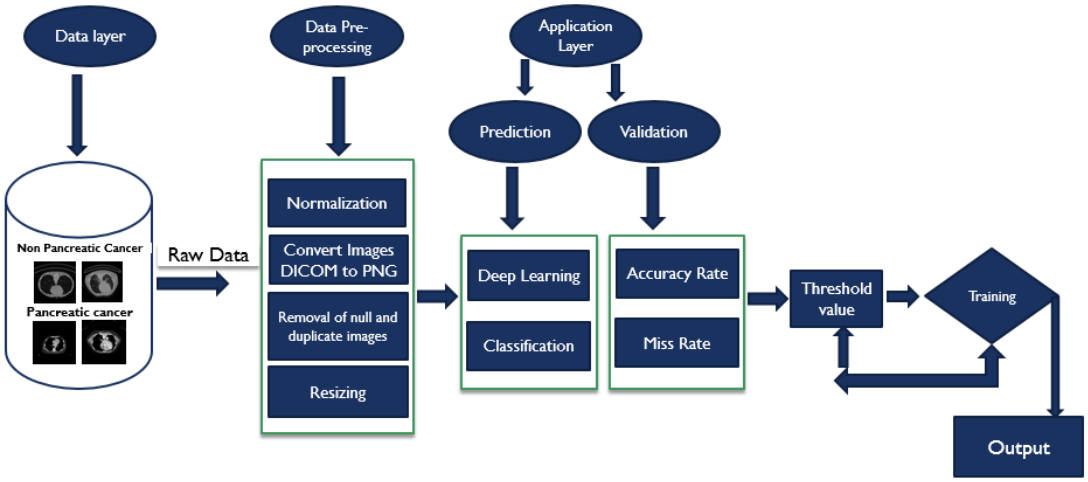
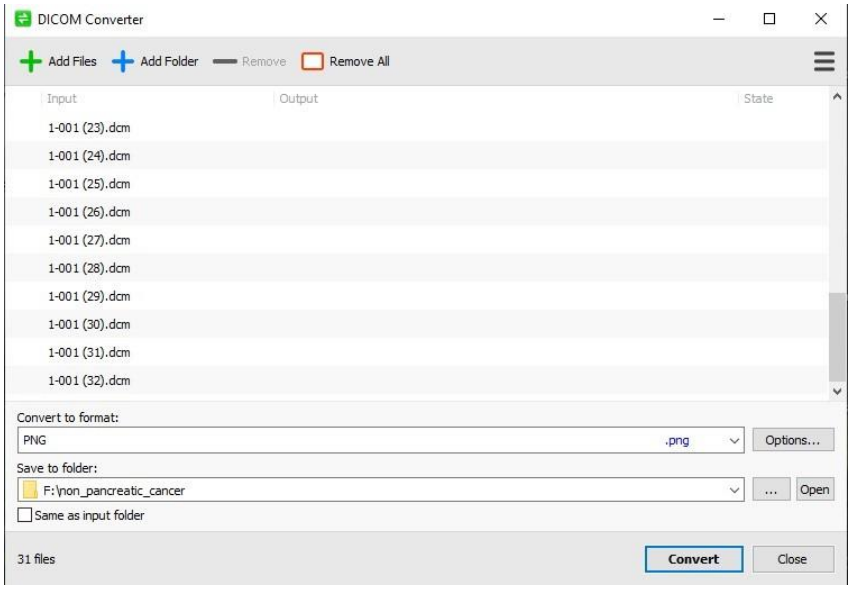


Figure 3: Architectural diagram of pancreatic cancer

Data Acquisition and Pre-processing: The image samples of pancreatic cancer once acquired then moves towards data pre-processing through raw data, the data pre-processing converts DICOM images to PNG format and then remove all null and duplicate images were removed manually. The dataset acquired images were almost 20300 in total. Dicom converter software was used to convert images from DICOM to PNG. DICOM converter is a free and offline software for images conversion DICOM to PNG or different images format, software was downloaded from the official DICOM Converter website then after installation the software was ready to use. At least 70-150 images at a time were inserted into the software and then were converted into PNG format until all the images were completed. Figure 11 shows the software interface of the DICOM to PNG converted tool used. After the removal of duplication and null images the final benchmark dataset consists of 20013 images for both pancreatic and non-pancreatic images. The PNG images were gone through a few image enhancement and pre-processing operations to enhance the images into a similar format here normalization, resizing were done for image enhancement. Figure 12 shows a few input images of the benchmark dataset after pre-processing.

SIMULATION AND RESULTS



Classification

Figure 4: DICOM to PNG converter.

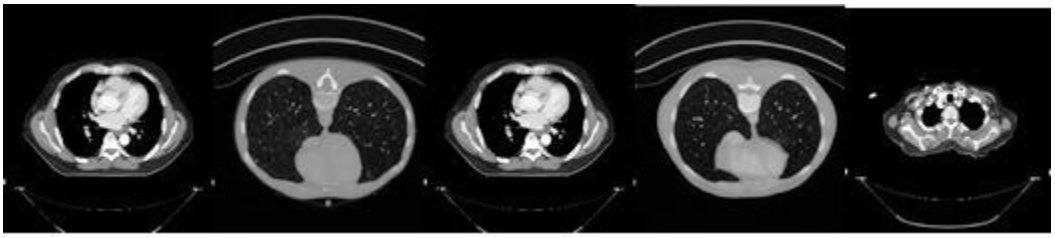


Figure 5: Input Dataset Images of CT Scan pancreatic and Non-Pancreatic cancer images

The final benchmark dataset, which comprises 20013 pictures of the two classes of pancreatic and non-pancreatic cancer, is created after pre-processing. The model makes use of a VGG16 model, which is regarded as one of the top vision model architectures on the market right now. The most notable characteristic of VGG16 is that they focused on having 3x3 filter convolution layers with a stride 1 instead than having a large number of hyper-parameters. They also consistently used the same padding and maxpool layer of 2x2 filter stride 2. The convolution and max pool layers are positioned consistently throughout the architecture. It has a SoftMax for output after two FC (fully connected layers) at the end. The number 16 in VGG16 refers to the fact that it

has sixteen layers of varying weights. With an estimated 138 million parameters, this network is fairly large.

The dataset statistics where the total number of input images are 20014 representing 2 classes, here 18013 images are used for training while 2001 image are used for validation. Here the training and validation process are carried out simultaneously with a total number of 30 epochs. Here the train-text split ratio is carried out to be 80-20% where 80% of the data is used for training and 20% is used for validation. Once the training and validation process is completed, the performance metrics are evaluated based on the learning experience of the model. The training and validation accuracy are evaluated with respect to the training and validation loss. moreover, precision, recall, f1-score, and support are evaluation in terms of training and validation processes. There is a certain threshold value that is implemented in the model here, a threshold value set is 90%. If the training accuracy fails to exceed the threshold function, then the training process will be carried out again.

The simulation and results are mentioned that states all the steps after the application of the methodology and architecture. The proposed system has been developed using python 3.7 while the tool for the development of the system is visual studio code. The model uses a custom VGG16 architecture using pytorch and keras libraries and functionalities. Figure 11 shows the number of classes used in the proposed methodology class "0" represents non pancreatic cancer while class "1" represents pancreatic cancer. The code uses a special function that checks the dataset images if they contain any improper images if so, the function returns the improper image count value with respect to images names in this case as the images were pre- processed so there were no bad or improper images, so the length is 0. Figure 12 shows the dataset statistics where the total number of input images are 20014 representing 2 classes, here 18013 images were used for training while 2001 images were used for validation that makes the 80-20 split.

```
processing class directory  non_pancreatic_cancer
processing class directory  pancreatic_cancer
 no improper image files were found
```

```
len(bad_file_list)
```

```
0
```

Figure 6: Number of classes used with no bad images

```
Found 20014 files belonging to 2 classes.
Using 18013 files for training.
Found 20014 files belonging to 2 classes.
Using 2001 files for validation.
Image -->  (32, 227, 227, 3) Label -->  (32,)
```

Figure 7: Statistics of Input Dataset

```
class_names = train_ds.class_names
print(class_names)
```

```
['non_pancreatic_cancer', 'pancreatic_cancer']
```

Figure 8: Names of Classes used in the dataset

Figure 6 displays the class labels of the input training dataset that involves two classes non- pancreatic cancer that represents 0 while the pancreatic cancer class represents 1. Once the dataset is loaded then the code uses a function to specifically pick random images of the dataset and display them as a random image matrix. Figure 14 shows the random image matrix of the input dataset of both classes.

Figure 8 explains the applied VGG16 model using keras, the model architecture is based upon convolution layers, pooling layers, and normalization layers. Here relu is used as an activation function for the hidden layers and SoftMax activation function is used in the output layers. As the model is based on the binary classification problem so SoftMax is the best option for the model. Figure 16 explains the training and validation

process in term of epoch the method utilizes a few 30 epoch where the training and validation processes are carried out parallelly. The training and validation accuracy at the last epoch is 100% and 100% respectively.

Figure 10 states the training and validation graph once the training and process is completed here the graph plot the training and validation accuracy with respect to each epoch while the training and validation loss is also shown with respect to every epoch. The training accuracy at epoch 1 started from 0.9984 and went to 100% at epoch 30 while the validation accuracy is shown to be 100% from epoch 1 till 30. The training loss is shown to be 0.007 at epoch1 and went to 0 at epoch 30 while the validation loss is 0 starting from epoch 1 to 30. Figure 11 shows the overall model accuracy after the training and validation process is completed.

Figure 11 states the confusion matrix of the training process, the training data consisted of 18013 images where 9010 images represent class 0 while 9003 images represent class 1. The model correctly trained on all input images resulting in a 100% training accuracy. Figure 13 states the confusion matrix for the validation process here the validation process uses 2001 images where 1029 images represent class 0 while 972 images represent class 1. Here the model correctly classified both classes with 100% where not even a single image is incorrectly classified. Figure 14 shows the class wise accuracy of the training process of the model, here the precision, recall, f1-score, and support of both classes. Moreover, the micro avg, macro avg and weighted avg are also shown. Figure 15 shows the class wise accuracy of the validation process of the model, here the precision, recall, f1-score, and support of both classes. Further, the micro avg, macro avg and weighted avg are also shown.

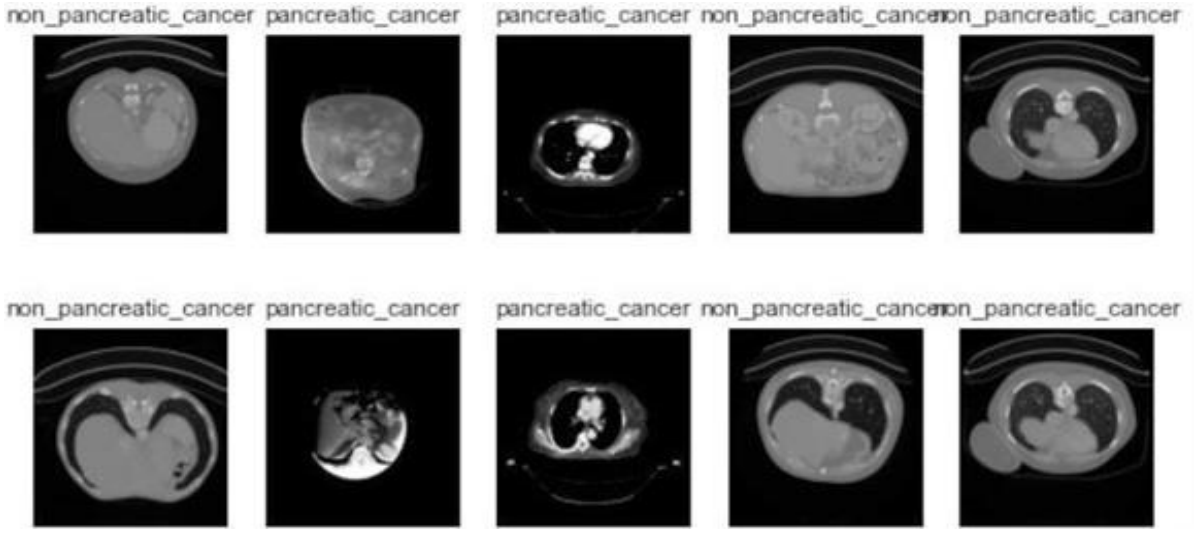


Figure 9: Image Matrix plot of training images used

```
model = keras.models.Sequential([
    keras.layers.Conv2D(filters=96, kernel_size=(11,11), strides=(4,4), activation='relu', input_shape=(227,227,3)),
    keras.layers.BatchNormalization(),
    keras.layers.MaxPool2D(pool_size=(3,3), strides=(2,2)),
    keras.layers.Conv2D(filters=256, kernel_size=(5,5), strides=(1,1), activation='relu', padding="same"),
    keras.layers.BatchNormalization(),
    keras.layers.MaxPool2D(pool_size=(3,3), strides=(2,2)),
    keras.layers.Conv2D(filters=384, kernel_size=(3,3), strides=(1,1), activation='relu', padding="same"),
    keras.layers.BatchNormalization(),
    keras.layers.Conv2D(filters=384, kernel_size=(3,3), strides=(1,1), activation='relu', padding="same"),
    keras.layers.BatchNormalization(),
    keras.layers.Conv2D(filters=256, kernel_size=(3,3), strides=(1,1), activation='relu', padding="same"),
    keras.layers.BatchNormalization(),
    keras.layers.MaxPool2D(pool_size=(3,3), strides=(2,2)),
    keras.layers.Flatten(),
    keras.layers.Dense(4096, activation='relu'),
    keras.layers.Dropout(0.5),
    keras.layers.Dense(4096, activation='relu'),
    keras.layers.Dropout(0.5),
    keras.layers.Dense(2, activation='softmax')
])
```

Figure 10: Proposed VGG16 architecture used

```
563/563 [==============================] - 34s 52ms/step - loss: 0.0072 - accuracy: 0.9984 - val_loss: 0.0000e+00 - val_accuracy: 1.0000
Epoch 2/30
563/563 [==============================] - 21s 38ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 0.0000e+00 - val_accuracy: 1.0000
Epoch 3/30
563/563 [==============================] - 21s 37ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 0.0000e+00 - val_accuracy: 1.0000
Epoch 4/30
563/563 [==============================] - 21s 37ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 0.0000e+00 - val_accuracy: 1.0000
Epoch 5/30
563/563 [==============================] - 21s 37ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 0.0000e+00 - val_accuracy: 1.0000
Epoch 6/30
563/563 [==============================] - 21s 37ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 0.0000e+00 - val_accuracy: 1.0000
Epoch 7/30
563/563 [==============================] - 21s 37ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 0.0000e+00 - val_accuracy: 1.0000
Epoch 8/30
563/563 [==============================] - 21s 37ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 0.0000e+00 - val_accuracy: 1.0000
Epoch 9/30
563/563 [==============================] - 21s 37ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 0.0000e+00 - val_accuracy: 1.0000
…
Epoch 29/30
563/563 [==============================] - 22s 39ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 0.0000e+00 - val_accuracy: 1.0000
Epoch 30/30
563/563 [==============================] - 21s 38ms/step - loss: 0.0000e+00 - accuracy: 1.0000 - val_loss: 0.0000e+00 - val_accuracy: 1.0000
```

Figure 11: Training and Validation statistics per epoch.

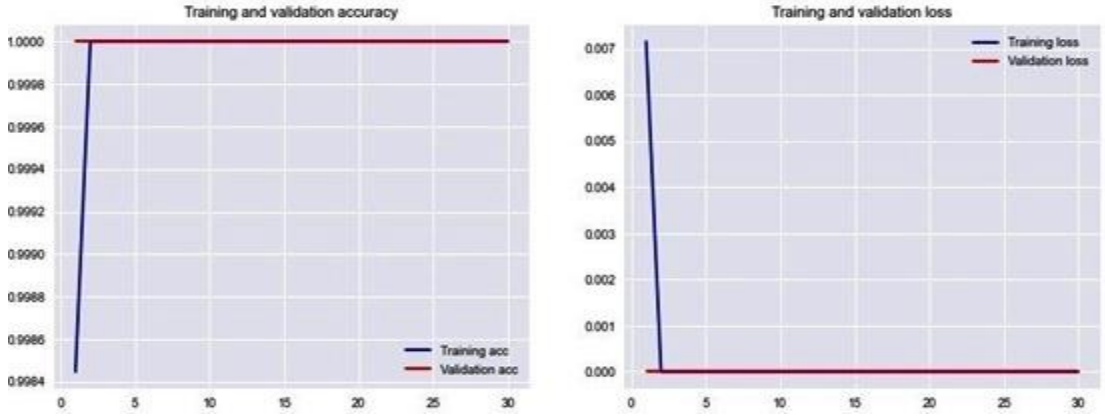


Figure 12: Training and Validation graph

Figure 13: Overall Validation accuracy of the proposed deep learning model

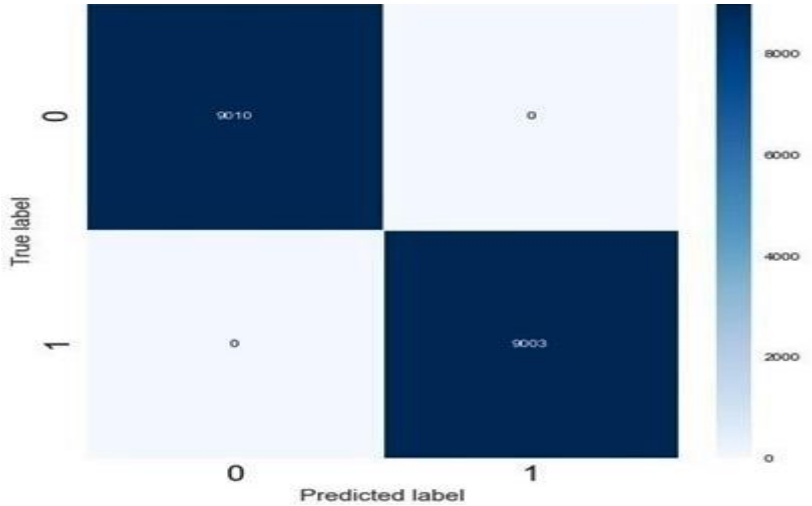


Figure 14: Confusion matrix of the training process

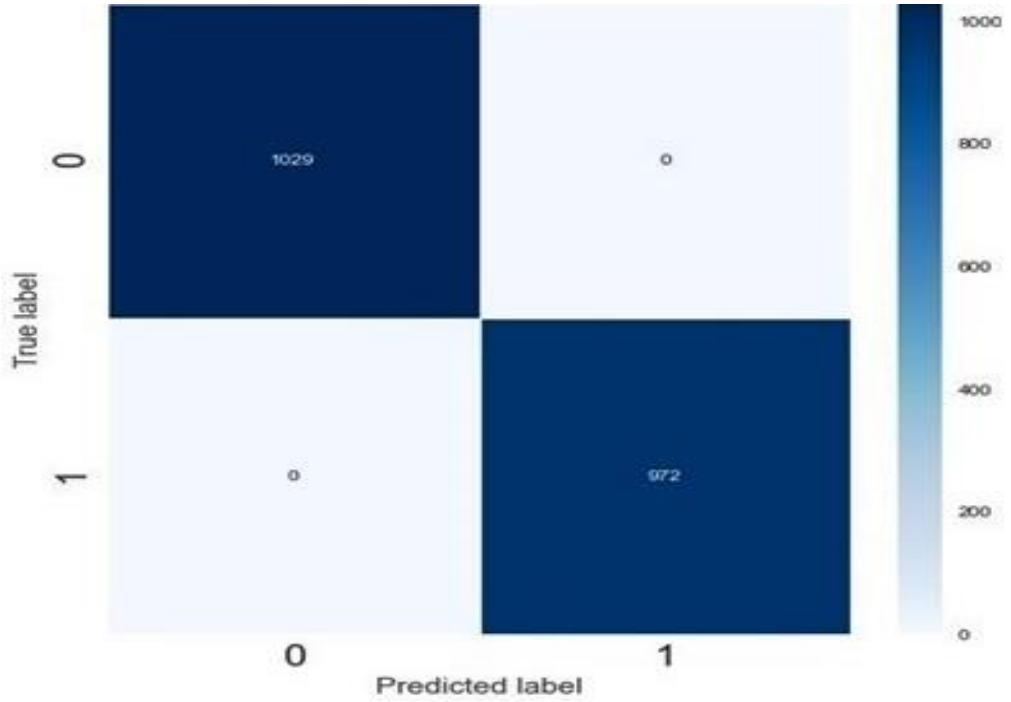


Figure 15: Confusion Matrix of the validation process

	precision	recall	f1-score	support
0	1.00	1.00	1.00	9010
1	1.00	1.00	1.00	9003
micro avg	1.00	1.00	1.00	18013
macro avg	1.00	1.00	1.00	18013
weighted avg	1.00	1.00	1.00	18013

Figure 16: Training accuracy per class

********** Test Data **********

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1029
1	1.00	1.00	1.00	972
micro avg	1.00	1.00	1.00	2001
macro avg	1.00	1.00	1.00	2001
weighted avg	1.00	1.00	1.00	2001

Figure 17: Validation accuracy per class

Table 1 states the detailed performance comparison of relative methodologies with the proposed methodology. Here various techniques of related authors have been cited. The proposed methodology uses a custom VGG16 model the accuracy of the proposed methodology surpassed the other existing techniques where the proposed methodology produced a validation.

Table 2: Performance comparison of proposed methodology with related work done

Author name and Year	Title of Research	Model	Dataset	Accuracy
Ali Al-Fatlawi., 2021	Deep learning Improves pancreatic cancer diagnosis using RNA based variants	CA19-9, deep learning	B4GALT5 and GSDMD	96%
Davide Placido., 2021	Pancreatic cancer risk prediction from disease trajectories using deep learning.	Machine Learning, Artificial Intelligence GRU, MLP	DNPR	95%
Seiya Yokoyama., 2020	Predicted Prognosis of Patients with pancreatic cancer by Machine Learning	Machine Learning	SVM and NNET	95%
Xiaodong Li, Peng Gao., 2020	A deep learning-based prediction of pancreatic adenocarcinoma with electronic health records	Deep Learning	EHR dataset	91.20%
Jonas cicenas.,	KRAS, TP53, CDKN2A, SMAD4, BRCA1, and BRCA2 Mutation in Pancreatic cancer	ML, AI	PI3K-AKT, PLC-PKC, and RAL	97%
Sabah Khudhair Abbas.,	Novel computer aided diagnostic system using synergic deep learning technique for early detection of pancreatic cancer	MATLAB, Deep learning	PCCD	99.23%
Proposed Methodology		**VGG16**	**Pancreatic and Non-pancreatic CT scan image dataset**	**100%**

Conclusion and future work: The in-depth study of medical health systems is giving computer systems a lot of chances to develop the most innovative innovations. More effective medical system implementations, including automated identification of health-related problems, are being made possible by these developments. The most important medical research is being done to predict cancer, which may harm many parts of the body and take many different forms. One of the most prevalent tumours that is predicted to be incurable is pancreatic cancer. Since it is located in the abdomen behind the stomach, it is often found to be unexpected and cannot be effectively treated once discovered. Consequently, the development of automated systems that can identify the stages of cancer and, if found, provide better diagnosis and treatment is being facilitated by advancements in medical research. One such field that has extended its study into medical imaging is deep learning, which, when combined with technology like CT/PET scan systems, automates the process of identifying patient issues. In order to predict the percentage of cancer spread in the pancreas using the threshold parameters as markers, the Convolutional Neural Network (CNN) model is utilised in this paper. The CNN model is embedded with the Gaussian Mixture model with EM algorithm to predict the key features from the CT scan. The suggested approach employs a proprietary deep learning algorithm that uses a VGG16 model to classify pancreatic cancer. The experiment is carried out using a dataset of CT scan images of the pancreas collected from the Cancer Image Archive TCIA, which has 20,013 of two classifications. The model outperformed other relevant current methodologies, producing a 100% training and validation accuracy.

REFERANCES

[1] K. L. Liu *et al.*, "Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation," *Lancet Digit. Heal.*, vol. 2, no. 6, pp. e303–e313, 2020, doi: 10.1016/S2589-7500(20)30078-9.

[2] B. Alizadeh Savareh *et al.*, "A machine learning approach identified a diagnostic model for pancreatic cancer through using circulating

microRNA signatures," *Pancreatology*, vol. 20, no. 6, pp. 1195–1204, 2020, doi: 10.1016/j.pan.2020.07.399.

[3] K. Sekaran, P. Chandana, N. M. Krishna, and S. Kadry, "Deep learning convolutional neural network (CNN) With Gaussian mixture model for predicting pancreatic cancer," *Multimed. Tools Appl.*, vol. 79, no. 15–16, pp. 10233–10247, 2020, doi: 10.1007/s11042-019-7419-5.

[4] M. M. Althobaiti, A. Almulihi, A. A. Ashour, R. F. Mansour, and D. Gupta, "Design of Optimal Deep Learning-Based Pancreatic Tumor and Nontumor Classification Model Using Computed Tomography Scans," *J. Healthc. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/2872461.

[5] G. Suman, A. Panda, P. Korfiatis, and A. H. Goenka, "Convolutional neural network for the detection of pancreatic cancer on CT scans," *Lancet Digit. Heal.*, vol. 2, no. 9, p. e453, 2020, doi: 10.1016/S2589-7500(20)30190-4.

[6] A. Al-fatlawi *et al.*, "RNA-Based Variants," 2021.

[7] J. M. Wyse and A. V. Sahai, "Endoscopic Ultrasound-Guided Management of Pain in Chronic Pancreatitis and Pancreatic Cancer: an Update," *Curr. Treat. Options Gastroenterol.*, vol. 16, no. 4, pp. 417–427, 2018, doi: 10.1007/s11938-018-0193-z.

[8] S. Walczak and V. Velanovich, "Improving prognosis and reducing decision regret for pancreatic cancer treatment using artificial neural networks," *Decis. Support Syst.*, vol. 106, pp. 110–118, 2018, doi: 10.1016/j.dss.2017.12.007.

[9] F. J. Harmsen, D. Domagk, C. F. Dietrich, and M. Hocke, "Discriminating chronic pancreatitis from pancreatic cancer: Contrast-enhanced EUS and multidetector computed tomography in direct comparison," *Endosc. Ultrasound*, vol. 7, no. 6, pp. 395–403, 2018, doi: 10.4103/eus.eus_24_18.

[10] K. Guo, W. Song, Y. Gong, S. Hu, W. Zhong, and W. Qiu, "Down-Regulation of Survivin Expression by siRNA Suppresses Proliferation and Enhances Chemosensitivity in Human Pancreatic Cancer Cell Line Panc-1," *Proc. - 2015 7th Int. Conf. Meas. Technol. Mechatronics Autom. ICMTMA 2015*, pp. 400–402, 2015, doi: 10.1109/ICMTMA.2015.102.

[11] J. Noorbakhsh *et al.*, "Deep learning-based cross-classifications reveal conserved spatial behaviors within tumor histological images," *Nat. Commun.*, vol. 11, no. 1, pp. 1–14, 2020, doi: 10.1038/s41467-020-20030-5.

[12] S. Hussein, P. Kandel, C. W. Bolan, M. B. Wallace, and U. Bagci, "Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches," *IEEE Trans. Med. Imaging*, vol. 38, no. 8, pp. 1777–1787, 2019, doi: 10.1109/TMI.2019.2894349.

[13] M. Yalchin, A. M. Baker, T. A. Graham, and A. Hart, "Predicting colorectal cancer occurrence in ibd," *Cancers (Basel).*, vol. 13, no. 12, pp. 1–28, 2021, doi: 10.3390/cancers13122908.

[14] L. C. Chu and E. K. Fishman, "Deep learning for pancreatic cancer detection: current challenges and future strategies," *Lancet Digit. Heal.*, vol. 2, no. 6, pp. e271–e272, 2020, doi: 10.1016/S2589-7500(20)30105-9.

[15] W. Y. Chen *et al.*, "The timing of continuous renal replacement therapy initiation in sepsis-associated acute kidney injury in the intensive care unit: The CRTSAKI Study (Continuous RRT Timing in Sepsis-associated AKI in ICU): Study protocol for a multicentre, randomised cont," *BMJ Open*, vol. 11, no. 2, pp. 1–8, 2021, doi: 10.1136/bmjopen-2020-040718.

[16] H. Kondylakis *et al.*, "IManageCancer: Developing a Platform for Empowering Patients and Strengthening Self-Management in Cancer Diseases," *Proc. - IEEE Symp. Comput. Med. Syst.*, vol. 2017-June, pp. 755–760, 2017, doi: 10.1109/CBMS.2017.62.

[17] K. Si *et al.*, "Fully end-to-end deep-learning-based diagnosis of pancreatic tumors," *Theranostics*, vol. 11, no. 4, pp. 1982–1990, 2021, doi: 10.7150/thno.52508.

[18] H. Kim *et al.*, "Biomarker panel for the diagnosis of pancreatic ductal adenocarcinoma," *Cancers (Basel).*, vol. 12, no. 6, 2020, doi: 10.3390/cancers12061443.

[19] Y. Lv, Y. Wang, Y. Tan, W. Du, K. Liu, and H. Wang, "Pancreatic cancer biomarker detection using recursive feature elimination based on Support Vector Machine and large margin distribution machine," *2017 4th Int. Conf. Syst. Informatics, ICSAI 2017*, vol. 2018-January, no. Icsai, pp. 1450–1455, 2017, doi: 10.1109/ICSAI.2017.8248514.

[20] M. Igarashi, "Artificial Intelligence-assisted system development in gastrointestinal endoscopy and surgery," *25th Opto-Electronics Commun. Conf. OECC 2020*, pp. 2020–2022, 2020, doi: 10.1109/OECC48412.2020.9273702.

[21] Z. Zhang, S. Li, Z. Wang, and Y. Lu, "A Novel and Efficient Tumor Detection Framework for Pancreatic Cancer via CT Images," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2020-July, pp. 1160–1164, 2020, doi: 10.1109/EMBC44109.2020.9176172.

[22] M. Li *et al.*, "Computer-Aided Diagnosis and Staging of Pancreatic Cancer Based on CT Images," *IEEE Access*, vol. 8, pp. 141705–141718, 2020, doi: 10.1109/ACCESS.2020.3012967.

[23] B. Dhruv, N. Mittal, and M. Modi, "Early and precise detection of pancreatic tumor by hybrid approach with edge detection and artificial intelligence techniques," *EAI Endorsed Trans. Pervasive Heal. Technol.*, vol. 7, no. 28, pp. 1–17, 2021, doi: 10.4108/eai.31-5-2021.170009.

[24] H. Hayashi *et al.*, "Recent advances in artificial intelligence for pancreatic ductal adenocarcinoma," *World J. Gastroenterol.*, vol. 27, no. 43, pp. 7480–7496, 2021, doi: 10.3748/wjg.v27.i43.7480.

[25] Z. Quo *et al.*, "Deep LOGISMOS: Deep learning graph-based 3D segmentation of pancreatic tumors on CT scans," *Proc. - Int. Symp. Biomed. Imaging*, vol. 2018-April, no. Isbi, pp. 1230–1233, 2018, doi: 10.1109/ISBI.2018.8363793.

[26] W. Xuan and G. You, "Detection and diagnosis of pancreatic tumor using deep learning-based hierarchical convolutional neural network on the internet of medical things platform," *Futur. Gener. Comput. Syst.*, vol. 111, pp. 132–142, 2020, doi: 10.1016/j.future.2020.04.037.

[27] W. Muhammad *et al.*, "Pancreatic Cancer Prediction Through an Artificial Neural Network," *Front. Artif. Intell.*, vol. 2, no. May, pp. 1–10, 2019, doi: 10.3389/frai.2019.00002.