COMPUTER AIDED DIAGNOSIS OF PANCREATIC TUMOR BASED ON CT IMAGES

A PROJECT REPORT

Submitted by

KASANUR JYOTHIRADHITHYA [Reg No:RA2011003010416]

N. MD. ATAHAR HAJEE [Reg No: RA2011003010892]

Under the Guidance of

Dr. A. M. J. MUTHUKUMARAN

Assistant Professor, Department of Computing Technologies

in partial fulfillment of the requirements for the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTING TECHNOLOGIES

COLLEGE OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR– 603 203

NOVEMBER 2023

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR–603 203

BONAFIDE CERTIFICATE

Certified that 18CSP109L / I8CSP111L project report titled "**COMPUTER AIDED DIAGNOSIS OF PANCREATIC TUMOR BASED ON CT IMAGES**" is the bonafide work of **KASANUR JYOTHIRADHITHYA [RA2011003010416]** and **N. MD. ATAHAR HAJEE [RA2011003010892]** who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported here in does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion for this or any other candidate.

Dr. A. M. J. Muthukumaran
SUPERVISOR
Assistant Professor
Department of Computing Technologies

Dr. B. Muruganantham
PANEL HEAD
Associate Professor
Department of Computing Technologies

Dr. M. Pushpalatha
HEAD OF THE DEPARTMENT
Department of Computing Technologies

Department of Computing Technologies

SRM Institute of Science and Technology

Own Work Declaration Form

Degree/Course **:**B.Tech in Computer Science and Engineering

Student Names **:** KASANUR JYOTHIRADHITHYA, N. MD. ATAHAR HAJEE

Registration Number: RA2011003010416, RA2011003010892

Title of Work **:** COMPUTER DIAGNOSIS OF PANCREATIC TUMOR USING CT IMAGES.

I/We here by certify that this assessment compiles with the University's Rules and Regulations relating to Academic misconduct and plagiarism, as listed in the University Website, Regulations, and the Education Committee guidelines.

I / We confirm that all the work contained in this assessment is our own except where indicated, and that we have met the following conditions:

- Clearly references / listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web,etc.)
- Given the sources of all pictures, data etc that are not my own.
- Not made any use of the report(s) or essay(s) of any other student(s)either past or present
- Acknowledged in appropriate places any help that I have received from others (e.g fellow students, technicians, statisticians, external sources)
- Compiled with any other plagiarism criteria specified in the Course hand book / University website

I understand that any false claim for this work will be penalized in accordance with the University policies and regulations.

DECLARATION:
I am aware of and understand the University's policy on Academic misconduct and plagiarism and I certify that this assessment is my / our own work, except where indicated by referring, and that I have followed the good academic practices noted above. **Student 1 Signature:** **Student 2 Signature:** **Date:**
If you are working in a group, please write your registration numbers and sign with the date for every student in your group.

ACKNOWLEDGEMENT

We express our humble gratitude to **Dr. C. Muthamizhchelvan**, Vice-Chancellor, SRM Institute of Science and Technology, for the facilities extended for the project work and his continued support.

We extend our sincere thanks to Dean-CET, SRM Institute of Science and Technology, **Dr. T. V. Gopal**, for his invaluable support.

We wish to thank **Dr. Revathi Venkataraman**, Professor and Chairperson, School of Computing, SRM Institute of Science and Technology, for her support throughout the project work.

We are incredibly grateful to our Head of the Department, **Dr. M. Pushpalatha,** Professor, Department of Computing Technologies, SRM Institute of Science and Technology, for her suggestions and encouragement at all the stages of the project work.

We want to convey our thanks to our Project Coordinator, **Dr. GODFREY**, Members, **Dr. A. M. J. MUTHUKUMARAN**, **Dr. J. NITHYASHRI , Dr. S. SIVAKUMAR and Dr. C. JAYAVARTHINI** , Panel Head, **Dr. B. MURUGANANTHAM**, Associate Professor and Panel Members, **Dr. J. NITHYASHRI** Assistant Professor, **Dr. S. SIVAKUMAR** Assistant Professor and **Dr. C. JAYAVARTHINI** Assistant Professor, Department of Computing Technologies, SRM Institute of Science and Technology, for their inputs during the project reviews and support.

We register our immeasurable thanks to our Faculty Advisor, **Dr.P.VELMURUGAN** Assistant Professor, Department of Computing Technologies, SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to our guide, **Dr. A. M. J. MUTHUKUMARAN**, Associate

Professor, Department of Computing Technologies, SRM Institute of Science and Technology, for providing us with an opportunity to pursue our project under her mentorship. He provided us with the freedom and support to explore the research topics of our interest. His passion for solving problems and making a difference in the world has always been inspiring.

We sincerely thank all the staff and students of Computing Technologies Department, School of Computing, S.R.M Institute of Science and Technology, for their help during our project. Finally, we would like to thank our parents, family members, and friends for their unconditional love, constant support and encouragement.

KASANUR JYOTHIRADHITHYA [Reg. No: RA2011003010416]

N. MD. ATAHAR HAJEE [Reg. No: RA2011003010892]

ABSTRACT

Regular abdominal X-rays may not give as much specific information on the pancreas as CT scans. Your physician may conduct a specialized CT scan to check for pancreatic cancer. Deep learning, computer vision, and medical image processing have advanced significantly since the Convolutional Neural Network (CNN) was used for visual data analysis. Deep learning-based methods may be used for medical image analysis, including the identification of pancreatic tumors. One type of malignant tumor disease is the pancreatic tumor. These methods use massive medical image datasets—typically from CT scans—to train deep neural networks that can recognize minute patterns that point to malignancies. The abstract offers a novel method for employing for the early detection and characterization of pancreatic cancers, with an emphasis on utilizing Convolutional Neural Network (CNN) models for improved accuracy. The aggressive characteristic of pancreatic cancer means that early identification is essential to enhancing therapy options and patient prognosis. With the CNN model at its center, this CAD system integrates a wide range of image processing methods and deep learning algorithms. The method performs exceptionally well in identifying areas of interest (ROIs) within CT scans by utilizing CNNs. With regards to the specificity of pancreatic tumor identification, this method produces impressive results.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

CNN	Convolutional neural network
ReLu	Rectified linear unit
CT	Computed tomography
MRI	Magnetic resonance imaging
CAD	Computed aided diagnosis
AI	Artificial intelligence
ROI	Region of Interest

CHAPTER-1

INTRODUCTION

1.1 INTRODUCTION

The pancreas is a tiny organ that is buried deep within the human body, making identification extremely challenging. Moreover, the primary cause of cancer death is failing to schedule drastic surgery at the best possible moment.

However, because abnormal texture features are difficult to discern and CT image quality varies between different CT scanners or operators, manual diagnosis still requires highly experienced medical professionals. As a result, research into developing a reliable deep-learning-based method for precise pancreatic tumor identification is becoming increasingly necessary.

Deep learning, computer vision, and medical image processing have advanced significantly since the Convolutional Neural Network (CNN) was used for visual data analysis.

The fields of deep learning, computer vision, and medical image processing have developed tremendously since the days of visual data analysis with Convolutional Neural Networks (CNNs).

Although it is uncommon, pancreatic cancer is among the deadliest kinds of the disease. That's because symptoms typically don't appear until the cancer is advanced, making treatment challenging. When cells in your pancreas mutate (alter) and proliferate uncontrollably, a tumor is formed, which is known as pancreatic cancer.

1.2 PANREATIC TUMOR

The term "pancreatic tumor" refers to a broad range of growths that may develop inside the pancreas, a vital organ situated behind the stomach. These growths fall into one of two categories: benign or cancerous. Pancreatic adenocarcinoma, also known as pancreatic cancer, is the most common malignant form. The features, behavior, and possible consequences of pancreatic tumors can vary greatly, thus a precise diagnosis and suitable therapy are essential.

Abdominal pain, involuntary losing weight, no appetite, and bowel habits are some of the ambiguous symptoms linked to pancreatic tumors. It might be challenging to diagnose and treat the tumor early as these symptoms frequently do not appear until it has progressed to an advanced stage.

Medical imaging and pathological investigation are used in tandem to diagnose pancreatic cancers. To see and detect anomalies in the pancreas, common diagnostic techniques include CT, MRI, and endoscopic ultrasonography. A biopsy is frequently necessary for a conclusive diagnosis. During this procedure, a tissue sample from the pancreatic tumor is taken, and its kind and potential malignancy are assessed under a microscope.

The stage of a pancreatic tumor must be established after diagnosis in order to inform therapy choices. The staging procedure determines the degree of tumor dissemination inside the pancreas or to other organs. Staging offers information about the patient's prognosis and aids in choosing the best course of action.

Depending on the type, stage, and general health of the patient, there are different treatment options for pancreatic tumors. For tumors that are surgically removed, a pancreatectomy, may be an option. Chemotherapy and radiation therapy are frequently used to target and control pancreatic cancer. Furthermore, immunotherapies and targeted treatments are beginning to show promise as therapy options in certain circumstances. Treatment decisions are usually chosen following a careful assessment and taking into account each patient's unique situation.

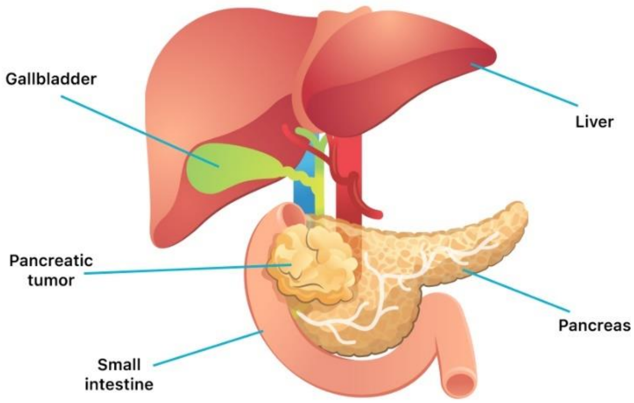

Fig.no: 1.1 PANCREATIC TUMOR

Due to late-stage diagnosis, the prognosis for pancreatic tumors, particularly pancreatic cancer, can be difficult to determine. The patient's overall health and the type and stage of the tumor all affect the prognosis. Given its reputation for being very aggressive, pancreatic cancer emphasizes the significance of early identification and timely treatments.

Chronic pancreatitis, smoking, a family history of pancreatic cancer, and specific genetic disorders are risk factors for the development of pancreatic tumors. A healthy weight, controlling chronic diseases, and giving up smoking are just a few examples of lifestyle modifications that can help lower the risk. Genetic counseling and screening may be taken into consideration for people with a family history of pancreatic cancer in order to identify cancers at an earlier, more manageable stage.

Receiving a diagnosis of pancreatic tumor can be physically and emotionally taxing. The assistance of loved ones, support groups, and medical experts is frequently beneficial to patients and their families. It is essential to provide patients with emotional and psychological support as they navigate their diagnosis and treatment plan.

Pancreatic tumor research is still being done in medicine. Researchers and medical experts are always trying to better understand the illness, create more potent therapies, and advance early detection techniques. For patients impacted by pancreatic cancers, improvements in prognosis and survival rates are possible because to developments in research and clinical trials.

1.3 CONVOLUTIONAL NEURAL NETWORT

The following factors make helpful for identifying pancreatic tumor:

Early Diagnosis: Pancreatic tumor can be accurately and early diagnosed with the use of CNN models, which evaluate medical images such as chest X-rays and CT scans. Early detection is crucial since prompt treatment can significantly improve patient outcomes.

Automation: By automating the detection process, CNN models eliminate the need for radiologists or other healthcare professionals to manually evaluate the data. This may lead to a quicker diagnosis and a decreased possibility of human error.

Efficiency: Due to its rapid processing of a large number of medical images, CNNs are helpful in hectic medical settings where speed and efficiency are essential.

Consistency: CNNs are able to produce consistent, dependable outcomes. The performance of a well-trained CNN is independent of factors such as weariness or expertise level, guaranteeing precise and reliable diagnosis of pancreatic tumor cases.

Support for Radiologists: By highlighting possible areas of concern, CNN models can be a very helpful tool for radiologists, helping them with case triage. This time savings may allow radiologists to concentrate on more complex cases.

Allocating Resources: By recognizing cases that require immediate attention and response, automated pancreatic tumor detection can help healthcare organizations allocate resources more effectively.

Quality Control: CNN models can aid in quality control by verifying radiologists' interpretations twice, reducing the possibility of a missed diagnosis.

Teaching Tool: Medical professionals can learn how to recognize pancreatic tumor in medical images and understand the associated patterns and features by using CNN models.

Research and Data Analysis: By analyzing a large dataset of medical images, CNN models can help with extensive medical research by helping researchers identify patterns, risk factors, and viable treatments for pancreatic tumor.
In telemedicine and remote healthcare settings, CNN-based detection systems can assist in the diagnosis of pancreatic tumor without the need for on-site radiologists. It's important to keep in mind that while CNNs might be useful tools for medical picture interpretation, their expertise shouldn't be substituted for that of a healthcare practitioner. The final say in diagnosis and treatment should always go to qualified medical professionals.

1.4 DEEP LEARNING

Preprocessing the photos includes scaling them to a constant resolution, normalizing the pixel values, and extending the dataset to enhance model generalization.

1.5 PROBLEM STATEMENT

The low survival rates of pancreatic cancer are largely due to late-stage diagnosis, which makes the disease extremely aggressive and frequently fatal. Improving patient outcomes and treatment options requires early diagnosis and thorough characterization of pancreatic cancers. However, it takes time and is prone to human error to manually interpret CT scans for tumor identification. A dependable Computer-Aided Diagnosis (CAD) system that uses deep learning—more specifically, Convolutional Neural Networks (CNNs)—is desperately needed in order to automatically identify and categorize pancreatic cancers in CT scans. The task at hand involves creating a computer-aided diagnostic (CAD) system that can quickly and accurately evaluate CT scans, identify regions of interest, and distinguish between benign and malignant lesions. This system would be an invaluable resource for early diagnosis and treatment planning.

1.6 OBJECTIVES

This project's primary goal is to develop and put into use a reliable Computer-Aided Diagnosis (CAD) system that uses Convolutional Neural Networks (CNNs) to accurately and early identify pancreatic cancers in CT images. For radiologists and clinicians, this CAD system will be a trustworthy tool that will aid in making an accurate diagnosis quickly.

1.7 PROJECT DOMAIN

The field of this project is the intersection of medical imaging and artificial intelligence (AI) in healthcare. It focuses on pancreatic tumors in CT images using Convolutional Neural Networks (CNNs). Several domains are involved in this initiative, which aims to help medical practitioners diagnose illnesses, particularly when it comes to identifying and describing pancreatic tumors using CT images.

1.8 SCOPE OF THE PROJECT

Developing a Computer-Aided Diagnosis (CAD) system that uses Convolutional Neural Networks (CNNs) to enable early and accurate pancreatic tumor detection in CT images is the primary goal of this project. With CNNs for enhanced accuracy, this CAD system will automatically detect regions of interest suggestive of malignancies and carry out sophisticated image analysis. The principal aim is developing a dependable instrument that augments the diagnostic proficiencies of medical practitioners, so culminating in preemptive measures and better patient consequences in instances of pancreatic cancer.

1.9 METHODOLOGY

This project's methodology is divided into multiple important stages. First and first, gathering data is essential. This entails obtaining a wide range of CT scans that show both benign and malignant pancreatic tumors. Working together with healthcare organizations guarantees ethical compliance and high-quality data. The CT scans are then ready for analysis through the further data preprocessing steps of noise reduction, contrast enhancement, and standardization.

Convolutional Neural Network (CNN) model creation and training are the project's central focus. This CNN does comprehensive picture analysis, identifying pertinent characteristics, and classifying. Optimizing the model allows for maximum performance. Concurrently, image segmentation techniques are applied to accurately delineate the pancreatic tumors in the areas.

Using a variety of datasets and cross-validation approaches, the performance of the CAD system is extensively checked during the crucial validation phase. Its diagnostic accuracy is evaluated using metrics including sensitivity, specificity, and accuracy. Collaboration with medical professionals is necessary since integration into clinical practice is a priority. Reports and documentation cover all aspect of the project, from algorithm descriptions to dataset details, with an emphasis on disseminating insightful information through thorough reports. The initiative is to equip medical practitioners with a trustworthy tool for early pancreatic tumor diagnosis through the use of this organized technique.

CHAPTER 2

LITERATURE REVIEW

2.1 LTERATURE REVIEW

An essential component of this research is a thorough review of the literature, which lays the groundwork for comprehending the most recent advancements in Convolutional Neural Networks (CNNs) and Computer-Aided Diagnosis (CAD) systems for pancreatic tumor diagnosis. In certain domains, a thorough literature study provides a solid theoretical framework for the project, allowing it to profit from the lessons learned and best practices described in previous research. This information influences the CAD system's development and execution, increasing its chances of success when it comes to early pancreatic tumor identification with CNNs and CT scans.

2.1.1 Computer aided diagnosis and staging of pancreatic cancer using CT images:

For preoperative PC diagnosis and staging, this work proposes a comprehensive medical computer-aided technique based on computed tomography (CT) images and an ensemble learning-support vector machine (EL-SVM). The least absolute shrinkage and selection operator (LASSO) algorithm was chosen for feature selection. The model optimization time decreased by 19.94 seconds without compromising accuracy when features were not chosen. The results of the experiment demonstrated that the classification accuracy for the following conditions was 86.61% for normal pancreas (normal)-pancreatic cancer early stage (early stage), 87.04% for normal-pancreatic cancer stage III (stage III), 91.63% for normal-pancreatic cancer stage IV (stage IV), 87.89% for normal-PC, 75.03% for early stage-stage III, 81.22% for early stages, and 82.48% for stage III-stage IV. The results of our tests show the feasibility and potential of our proposed method for clinical application in the preoperative diagnosis and staging of PC with CT scans.

2.1.2 Automatic Segmentation of Pancreas and Pancreatic Tumor: A Review of a Decade of Research:

This paper's primary goal is to provide an overview of the various automated methods for segmenting. Many indices, such as the Jaccard index (JI), recall, dice similarity coefficient (DSC), sensitivity (SI), specificity (SP), precision (Pr), and recall, will also be used in a comparison analysis. Lastly, a summary of the pancreas and tumor segmentation study limits and future research directions is provided.

2.1.3 Harmonic Motion Imaging for Abdominal Tumor Detection and High-Intensity Focused Ultrasound Ablation Monitoring: An In Vivo Feasibility Study in a Transgenic Mouse Model of Pancreatic Cancer:

The purpose of this study was to evaluate the feasibility of employing HMI for pancreatic tumor identification and high-intensity focused ultrasound (HIFU) therapy monitoring. The HMI system consisted of a focused ultrasound transducer that generated sinusoidal radiation force to induce oscillatory tissue motion at 50 Hz, and a diagnostic ultrasound transducer that used a 1-D cross-correlation algorithm to detect the axial tissue displacements based on acquired radio-frequency signals. To detect pancreatic cancers, HMI images were generated for both normal pancreases in wild-type mice and pancreatic tumors in transgenic animals. The obtained HMI images showed a clear distinction between cancerous and healthy pancreases, as well as an average peak-to-peak HMI displacement ratio.

2.1.4 Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches:

To enhance tumor characterization, we provide both supervised and unsupervised machine learning techniques in this work. We demonstrate significant gains in supervised learning with our first strategy, which uses a 3D Convolutional Neural Network and Transfer Learning, using deep learning techniques. Then, motivated by the radiologists' analyses of the scans, we show how to incorporate task-dependent feature representations into a CAD system using a graph-regularized sparse Multi-Task Learning (MTL) framework. The second option addresses the problem of labeled training data scarcity, which is a common problem in medical imaging applications, by exploring an unsupervised learning technique. We suggest using proportion-SVM to characterize tumors, drawing inspiration from computer vision techniques that utilize label proportion (LLP) learning. We also aim to get the fundamental answer to the question of how useful "deep features" are for unsupervised tumor classification. We evaluate our proposed unsupervised as well as supervised algorithms on two different tumor identification tasks utilizing 1018 CT as well as 171 MRI images for the lungs and pancreatic, respectively, and achieve the state-of-the-art sensitivity and specificity outcomes in each instance.

CHAPTER 3

PROJECT DESCRIPTION

3.1 EXISTING SYSTEM

For natural photos, a variety of techniques have been put forth to deal with noisy label classification. Describe a method for distributing weights to training samples that makes use of a separate, clean validation set. In order to optimize the gradient update, their intuition tells them to apply larger weights to clean training samples and smaller weights to noisy data. The invention and use of noisy label classification techniques in medical imaging data utilizing support vector machines with 96.7% accuracy has received comparatively little attention.

There are some many more ways like:

1. Medical History and Symptoms: Physicians begin by inquiring about the patient's past health, as well as any current symptoms, digestive problems, changes in bowel habits are common symptoms of pancreatic tumors.

2. Laparoscopy: Laparoscopic procedures can be used to obtain tissue samples for biopsies as well as to visually check the pancreas and surrounding tissues.

3. Physical Examination: A physical examination can assist medical professionals in detecting any apparent indications of abdominal tumors or jaundice.

4. Imaging Tests: To see the pancreas and any anomalies, a variety of imaging tests are performed. CT scans are one type of them.

3.2 LIMITATIONS IN EXISTING SYSTEMS

• Late Detection: Because signs of pancreatic cancer usually manifest late in the disease, patients are frequently discovered at an advanced stage. It is difficult to diagnose early with conventional procedures.

• Interobserver Variability: Different radiologists may interpret medical pictures differently, which could result in contradicting findings.

• Effectiveness: It might take a lot of time for medical personnel to review a lot of photos, which can cause delays in diagnosis.

• False Positives and Negatives: Using conventional techniques can lead to false positives, which mean a tumor was mistakenly identified, or false negatives, which mean a tumor was missed.

3.3 PROPOSED SYSTEM

• Using CT scan pictures to identify pancreatic tumors. After preprocessing these photos with image processing techniques, the tumor region in the image is classified using CNN model architecture.

• The CNN model architecture is used for the classification. In this, the classes indicated in the image are used to teach the system.

• Using contrast-enhanced CT scans of patients, we train a CNN to discriminate between healthy and pancreatic cancer.

3.4ADVANTAGES OF PROPOSED SYSTEM

• CNNs can provide many benefits when it comes to the identification of pancreatic cancers using medical images such as CT scans:

1. Feature extraction: Without the requirement for human feature engineering, CNNs automatically extract pertinent features from the input images. This is particularly helpful in medical imaging, where it might be simple to spot minor patterns.

2. Adaptation: Large datasets may be used to train CNNs, giving them the ability to learn from a variety of cases and adjust to differences in tumor appearance.

3. Early diagnosis: By identifying tumors at more curable stages, CNNs may be able to help in early tumor diagnosis, which could lead to better patient outcomes.

4. Consistency: Because CNNs are less prone to human error and weariness, tumor detection results may be more consistently achieved.

3.5 FEASIBILITY STUDY

A feasibility study is conducted to assess the project's viability and evaluate the advantages and disadvantages of the suggested solution. It is necessary to assess the use of masks in crowded settings. The feasibility study is done in three different ways.

1. Economic Viability
2. Empirical Viability
3. Practicality in Society

3.5.1 ECONOMIC FEASIBILITY

High-priced equipment is not needed for this suggested solution. Within the software now available, this project can be created.

3.5.2 TECHNICAL FEASIBILITY

The suggested system is a deep learning model in its entirety. Python is the process language utilized in this project, and the primary tools utilized. Tools can be obtained for free, and the technical know-how needed to use them is doable. We can infer that the project is technically doable from this.

3.5.3 SOCIAL FEASIBILITY

Determining a project's social feasibility is the first step in determining its acceptability. There are no social problems and our initiative is environmentally beneficial to society. Every member of society will be able to accept our project and take responsibility for the environment and society. The system has a very high degree of acceptability, and this is dependent on the techniques used. The society is well familiar with our system.

3.6 SYSTEM SPECIFICATION

3.6.1 HARDWARE REQUIREMENTS

- Intel i5-8250 CPU @1.60GHz 1.80GHz
- 512 GB SSD
- NVDIA GEFORCE
- CPU QUAD CORES

3.6.2 SOFTWARE REQUIREMENTS

- GOOGLE COLAB
- ANACONDS
- ANACONDA PROMPT
- PYTHON
- VISUAL STUDIO

CHAPTER 4

MODULE DESCRIPTION

4.1 MODULES

• Data set: An image dataset is a group of images used in applications to testing, validation, and also training.

• Data Pre-processing: Improving the quality of the photos, extracting pertinent data, and getting them ready for precise and efficient processing are the objectives of this step.

• Feature extraction: This method involves condensing and meaningfully representing the most significant aspects of the raw data in order to prepare it for additional analysis.

• CNN: A deep learning architecture called a convolutional neural network was created specifically for visual and spatial input.

4.1.1 DATA COLLECTION

The project's data gathering procedure is an essential stage that lays the groundwork for the creation and instruction of the system to detect pancreatic tumors. The data collecting phase is characterized by the following essential elements:

Dataset Choosing: The meticulous selection of an extensive and representative dataset of CT scans is the initial step.

Quality Assurance of Data: It is essential to guarantee the medical pictures' consistency and quality. As part of this, make sure the photos are as high-resolution as possible.

Size and Diversity of the Dataset: A sizable dataset comprising a broad range of instances is necessary for the efficient training and validation of the CAD system.

TUMOR DATA:

Fig.no: 4.1

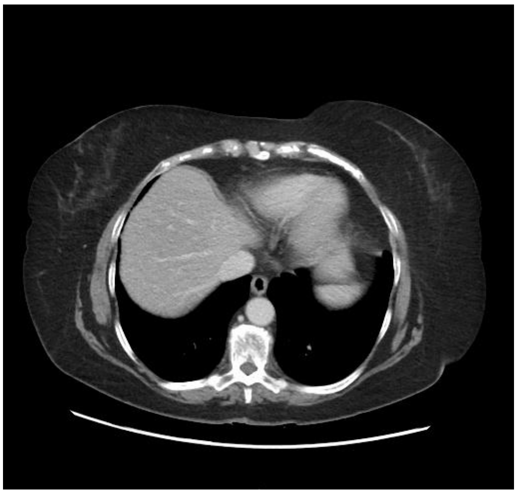

Fig.no: 4.2

NORMAL DATA:

Fig.no: 4.3

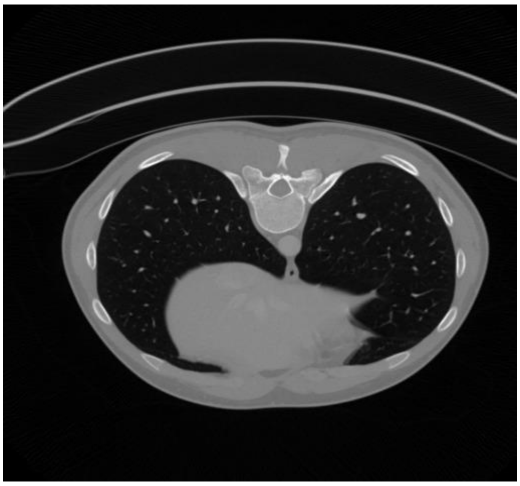

Fig.no: 4.4

4.1.2 DATA PRE-PROCESSING

When developing a system to detect pancreatic tumor in CT images, data preparation is an essential step. To make sure the data is in an appropriate format for analysis and model training, there are a few essential processes involved.

1. Data Augmentation: A variety of data augmentation methods are applied to the photos using the ImageDataGenerator. These methods consist of rescaling, horizontal flipping, shearing, and random zooming. Data augmentation helps to increase the dataset's variety. Data augmentation can assist the model in learning to handle changes in image orientation, scale, and other parameters in medical imaging applications.

2. Image Resizing: A uniform goal size of 224 by 224 pixels is used to resize the images. In order to guarantee that every image in the dataset has uniform dimensions—a prerequisite for training many machine learning models, especially CNNs—image scaling is frequently required.

3. Batch Generation: Pre-processed and enhanced photos are produced in batches by the code. In machine learning, batching is a popular technique because it reduces memory needs and facilitates efficient training at a time. In this instance, 32 image batches are created simultaneously.

4. Binary Classification Setup: The preprocessing code appears to be intended for a binary classification problem, as shown by the class_mode option being set to 'binary'. It might be used to categorize photos into two classifications, such as "tumor" and "non-tumor," or "malignant" and "benign," in the context of a project for pancreatic tumor identification.

4.1.3 FEATURE EXTRACTION

In this project, the Convolutional Neural Network (CNN) is the main tool for feature extraction. CNNs have the ability to recognize and extract pertinent features from photos automatically. These are components found in the pictures. Internal layers of the CNN automatically pick up and reflect these features during training.

4.1.4 CONVOLUTIONAL NEURAL NETWORK

One kind of Deep Learning neural network design that is frequently utilized in computer vision is the convolutional neural network (CNN). The branch of artificial intelligence known as "computer vision" gives computers the ability to comprehend and analyze images and other visual input. Although there are various types of neural networks, CNNs are the recommended network architecture for object identification and recognition in deep learning. They are therefore ideal for computer vision (CV) jobs and critical object recognition.

A deep learning CNN is composed of three distinct parts: the fully connected (FC) layer, the pooling layer, and the convolutional layer. The convolutional layer comes before and ends up before the FC layer.

The CNN gets more sophisticated as it moves from the convolutional layer to the FC layer. Because of its escalating intricacy, the CNN is able to recognize increasingly significant areas and intricate details of an image until ultimately identifying the item in its whole.

4.2 CNN LAYERS:

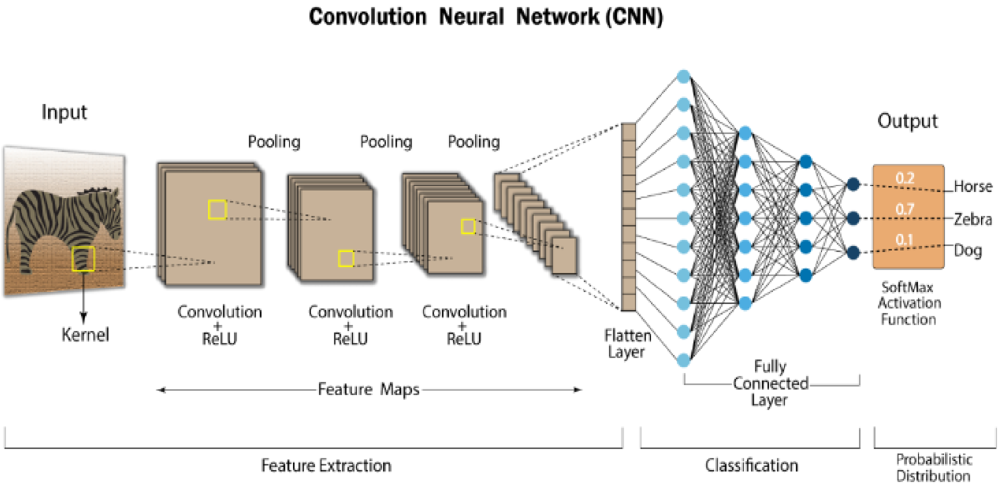

Fig.no: 4.5 LAYERS

4.2.1 FEATURE EXTRACTION LAYERS

Typical CNN feature learning layers include convolutional, max-pooling, and possibly additional normalization layers. A brief synopsis of each is provided below:

Convolutional Layers: These layers are used to recognize elements or trends in the source pictures, including edges, materials, and more complex patterns, using learnable filters or kernels. Convolutional layers are the essential parts of a CNN that are utilized to extract features.

Max-Pooling Layers: The use of max-pooling in convolutional layers results in feature maps with a lower spatial dimension. It comprises selecting the greatest value from a range of values (often a 2x2 grid) by down sampling the feature maps. This procedure helps to retain the most important information while reducing computer complexity.

Feature learning layers use pooling and convolution to extract relevant information from the input images. Because these layers are taught to recognize significant patterns and structures, they are essential to your model's capacity to detect pneumothorax. By transferring the learned data to subsequent layers for categorization, the model is ultimately able to generate remarkably accurate predictions.

Convolutional + Activation(ReLu) layers:

These layers apply convolution operations to the input data. Convolution is the technique of using a set of learnable filters, also called kernels, to perform element-wise multiplications and summations on an input image. Among other things, this process helps identify the image's edges, textures, and more complex patterns.

After every convolution step, an activation function is added to the network to introduce non-linearity. ReLU is best activation function. The positive values in the characteristic maps remain unaltered, but any negative ones are turned to zeros. This introduces non-linearity into the data, allowing the network to find complex connections within it:

$\max(0,x) = f(x)$

In the context of pneumothorax identification utilized to improve properties from photos. As network processes the data across multiple layers, it gets the capacity to recognize patterns, edges, and other relevant elements in the images. The ReLU activation function was included to allow correlations.

Convolution + ReLU levels are essential for your CNN model to be able to recognize tumors in clinical images because they allow the network to become more irregular and facilitate the extraction of features.

Pooling layers:

In a CNN architecture, Max-pooling is the most popular kind of pooling. This is how it operates:

Max-Pooling: In max-pooling, the maximum value within a sliding window (often 2x2 or 3x3) is chosen at each step as it progresses across the feature map. The feature map that has been down sampled then keeps this maximum value. Max-pooling aids in reducing the feature map's spatial dimensions while keeping the most noticeable features. It helps with translation invariance in particular.

The following is a definition of max-pooling:

for each and every value in the pool area

Maximum-Pooling(x) equals maximum(x) in Equation (1) for every value within the pool area.

Average Pooling: Average pooling is another popular pooling method. Rather than choosing the highest value within the pool area, it calculates the mean of all values. Average pooling is less likely to exclude important data and can assist lessen overfitting.

Goal: Reducing feature maps' size without sacrificing their most crucial features is the main goal of pooling layers. This decrease in spatial dimensions is beneficial in a number of ways:

It lessens the network's deeper levels' computing strain.
It works by lowering the number of parameters, which helps prevent overfitting.

Because of the degree of translation invariance it offers, the network can identify characteristics in the input image regardless of where they are precisely located.

4.2.2 CLASSIFIATION LAYERS

Flatten Layer:

An essential part of a CNN is the flatten layer, which is usually applied immediately before the fully linked (dense) layers.
Convolutional and pooling layers provide multi-dimensional feature maps or tensors, which it's main job is to reshape into a one-dimensional vector. The fully connected layers receive this one-dimensional vector after which it is used for tasks involving regression or classification.
Essentially, the Flatten layer transforms feature maps' spatial structure into a linear format.
The Flatten layer, for instance, would change a 2D feature map with 16 elements that is 4 by 4 in size into a 1D vector with 16 elements if the output of the pooling layer was that type of data.

Fully connected(Dense) layer:

The classic neural network layer known as a Fully Connected layer has all of its neurons connected to all of the neurons in the layers above and below.

After the convolutional and pooling layers have collected characteristics from the input data, these layers are usually found toward the end of the network. Within a Dense layer, every neuron calculates.

Fully Connected layers are probably utilized in your particular project for the last classification or detection task. They create the final output by flattening the feature vectors, which are the result of earlier layers.
In your project, the fully connected layers provide the function of carrying out the real classification or detection by using the features that the convolutional and pooling layers retrieved earlier. These layers are in charge of deriving the final predictions and figuring out the links and patterns in the feature representations.

Based on the acquired features and patterns, they might be able to assist you in determining whether or not a chest X-ray image shows indications of a tumor.

4.3 PROCESSING STAGES

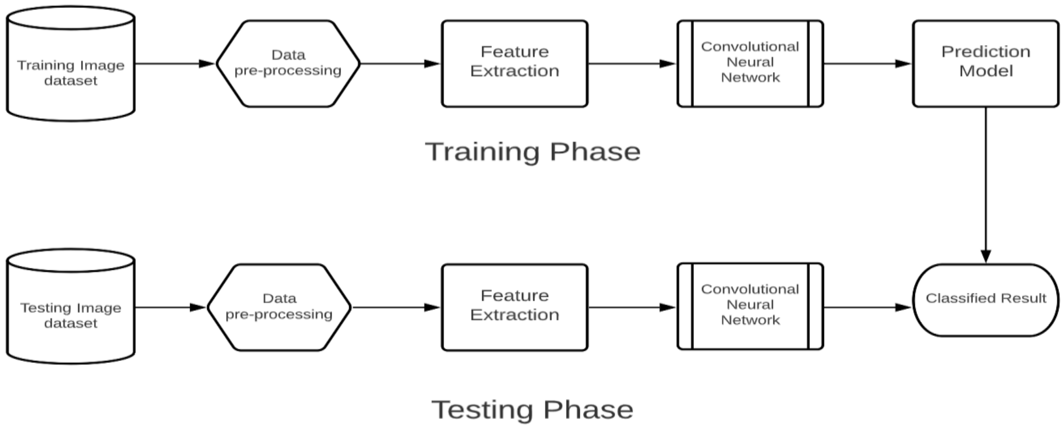

Fig.no: 4.6 STAGES

1. Training Image Dataset:

• A variety of CT pictures with malignant pancreatic tumors are gathered to create a training image dataset.

• The dataset needs to be properly annotated in order to provide training ground truth for the model.

2. Data Pre-processing:

• Preparing the dataset for training and making sure it is consistent, high-quality, and compatible with the CNN model is known as data pre-processing.

• Pre-processing processes could involve applying data augmentation techniques to boost dataset diversity, noise reduction, contrast enhancement, image scaling, normalization, and so on.

• Patient privacy is safeguarded by addressing ethical considerations and data anonymization.

3. Feature Extraction:

• For feature extraction, convolutional neural networks, or CNNs, are employed.

• From the previously processed images, the convolutional layers of the CNN automatically identify and extract pertinent features. These characteristics stand for forms, textures, and patterns typical to pancreatic tumors.

4. Convolutional Neural Network (CNN):

• For the project, a CNN architecture is created, accounting for variables such as filter sizes, layer count, and network depth.

• Using the pre-processed dataset, CNN is trained. The model learns to relate picture features to tumor classifications by processing the features retrieved from the images through the layers of the network.

• Transfer learning can be used to fine-tune pretrained models for the particular job of pancreatic tumor identification, using them as a starting point.

5. Prediction Model:

• The prediction model is the trained CNN model.

• It gives output predictions, usually indicating whether the input image contains a benign or malignant tumor, given pre-processed and feature-extracted images as input.

6. Classified Result:

• The input photos are classified by the prediction model, which generates findings. Radiologists and physicians can use the classified results to assist determine whether pancreatic cancers are present while making diagnostic judgments.

CHAPTER 5
IMPLEMENTATION

5.1 CODE

```
!unzip /content/test.zip
import warnings
warnings.filterwarnings('ignore')
import numpy as np
import matplotlib.pyplot as plt
import os
import math
import shutil
import glob
#count the number of images in the respective classes
ROOT_DIR="/content/trainval"
number_of_images={}
for dir in os.listdir(ROOT_DIR):
  number_of_images[dir]=len(os.listdir(os.path.join(ROOT_DIR,dir)))
number_of_images.items()
def dataFolder(p,split):
  #we creat ea training folder
  if not os.path.exists("./"+p):
    os.mkdir("./"+p)
    for dir in os.listdir(ROOT_DIR):
      os.makedirs("./"+p+"/"+dir)
      for img in
np.random.choice(a=os.listdir(os.path.join(ROOT_DIR,dir)),size=(math.
```

```
floor(split*number_of_images[dir])-5),replace=False):
        O=os.path.join(ROOT_DIR,dir,img)
        D=os.path.join("./"+p,dir)
        shutil.copy(O,D)
        os.remove(O)
  else:
    print("Train Folder exist")
dataFolder("trainn",0.85)
dataFolder("valn",0.15)
number_of_images={}
for dir in os.listdir(ROOT_DIR):
  number_of_images[dir]=len(os.listdir(os.path.join(ROOT_DIR,dir)))
number_of_images.items()
pip install Keras-Preprocessing
from keras.layers import Conv2D,MaxPool2D, Dropout, Flatten, Dense,
BatchNormalization, GlobalAvgPool2D
from keras.models import Sequential
from keras_preprocessing.image import ImageDataGenerator
import keras
model= Sequential()

model.add(Conv2D(filters = 16,
kernel_size=(3,3),activation='relu',input_shape=(224,224,3)))

model.add(Conv2D(filters = 36, kernel_size=(3,3),activation='relu'))
model.add(MaxPool2D(pool_size=(2,2)))
```

```
model.add(Conv2D(filters = 64, kernel_size=(3,3),activation='relu'))
model.add(MaxPool2D(pool_size=(2,2)))

model.add(Conv2D(filters = 128, kernel_size=(3,3),activation='relu'))
model.add(MaxPool2D(pool_size=(2,2)))

model.add(Dropout(rate=0.25))

model.add(Flatten())
model.add(Dense(units=64, activation='relu'))
model.add(Dropout(rate=0.25))
model.add(Dense(units=1,activation='sigmoid'))

model.summary()
model.compile(optimizer='adam',loss=keras.losses.binary_crossentropy,metrics=['accuracy'])
def preprocessingImages1(path):
  '''
  input: Path
  output: Pre processed images
  '''
  image_data= ImageDataGenerator(zoom_range=0.2,shear_range=0.2,rescale=1/255,horizontal_flip=True)
image=image_data.flow_from_directory(directory=path,target_size=(224,224),batch_size=32,class_mode='binary')
```

```
  return image
path='/content/trainn'
train_data=preprocessingImages1(path)
path='/content/test'
test_data=preprocessingImages1(path)
def preprocessingImages2(path):
  '''
  input: Path
  output: Pre processed images
  '''
  image_data= ImageDataGenerator(rescale=1/255)

image=image_data.flow_from_directory(directory=path,target_size=(224,22
4),batch_size=32,class_mode='binary')
  return image
path='/content/valn'
val_data=preprocessingImages2(path)
from keras.callbacks import ModelCheckpoint, EarlyStopping

#early stopping
es=EarlyStopping(monitor="val_accuracy",min_delta=0.01,patience=3,verb
ose=1,mode='auto')

#ModelCheckpoints
mc=ModelCheckpoint(monitor="val_accuracy",filepath="./bestmodel.h5",v
erbose=1,save_best_only=True,mode='auto')
```

```
cd=[es,mc]
#Model Training

hs=model.fit_generator(generator=train_data,
                       steps_per_epoch=8,
                       epochs=30,
                       verbose=1,
                       validation_data=val_data,
                       validation_steps=16,
                       callbacks=cd)
h=hs.history
h.keys()
import matplotlib.pyplot as plt
plt.plot(h['accuracy'])
plt.plot(h['loss'])
from keras.models import load_model
model=load_model("/content/bestmodel.h5")
#Model Accuracy
acc=model.evaluate_generator(test_data)[1]
print("Accuracy of our model is", acc)
from keras_preprocessing.image import load_img,img_to_array
path="/content/test/normal/1-001.jpg"
img=load_img(path,target_size=(224,224))
input_arr=img_to_array(img)/225
plt.imshow(input_arr)
plt.show()
```

```
input_arr=np.expand_dims(input_arr,axis=0)
'''predict_x=model.predict(input_arr)
classes_x=np.argmax(predict_x,axis=1)
pred=model.predict_classes(input_arr)[0][0]'''
pred=(model.predict(input_arr) > 0.5).astype("int32")
if(pred==0):
  print("The image is normal")
else:
  print("The image is having a Tumor")
```

5.2 OUTPUT

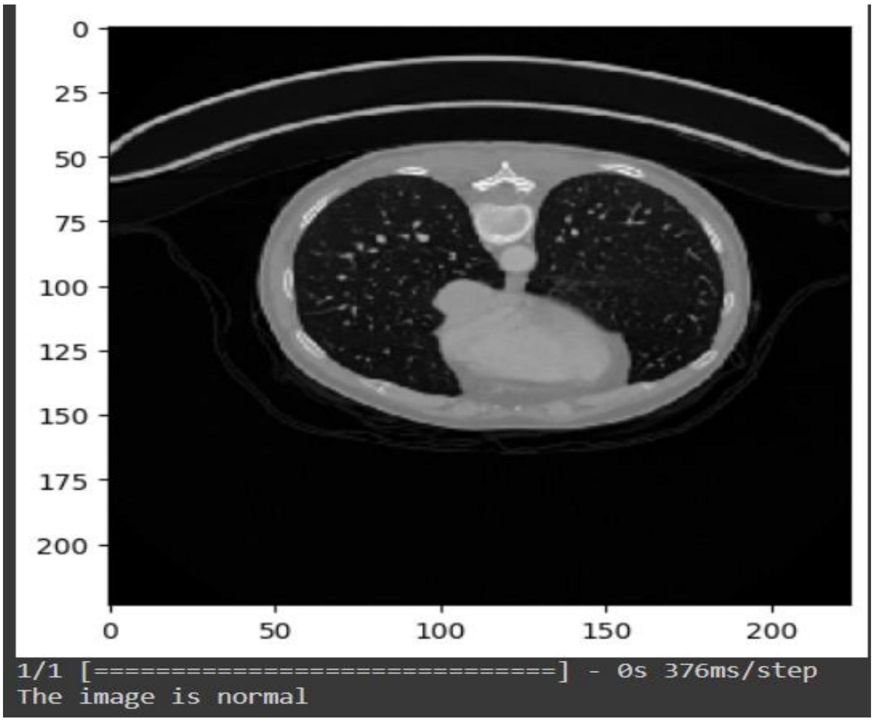

Fig.no: 5.1 OUTPUT

This output contains the identification of the tumor in the CT images.

5.3 ACCURACY

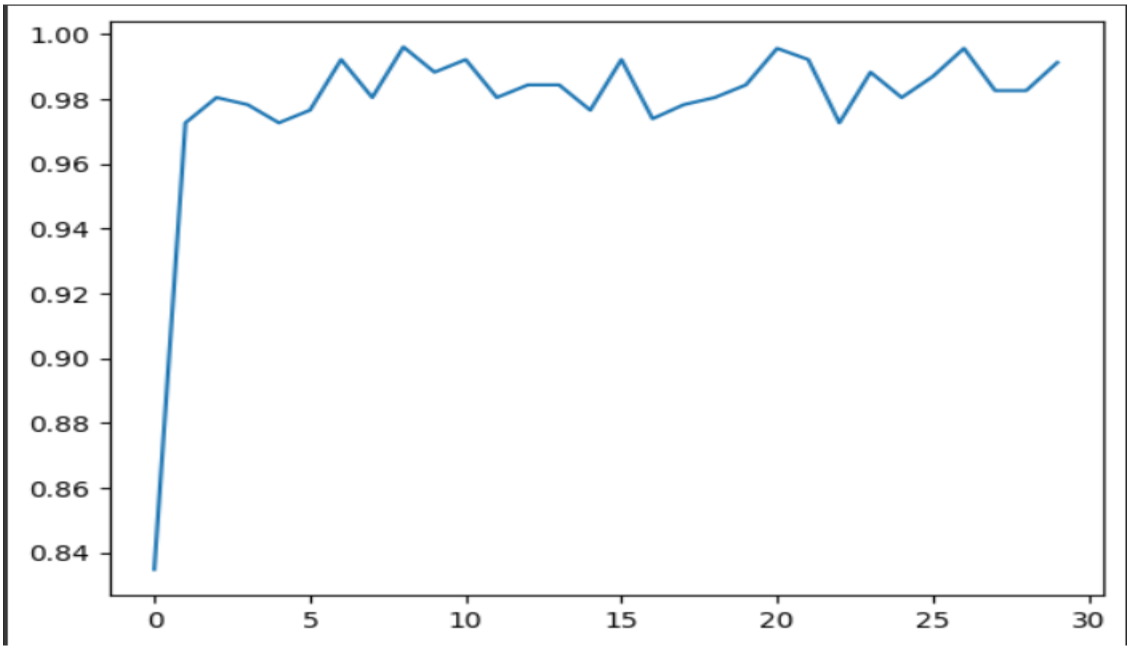

Fig.no: 5.2 ACCURACY GRAPH

```
WARNING:tensorflow:Can save best model only with val_accuracy available, skipping.
8/8 [==============================] - 5s 554ms/step - loss: 0.0310 - accuracy: 0.9913
```

Fig.no: 5.3 Accuracy

The training progress is indicated in this section of the message. It informs you about the training process, including the current loss and accuracy on the training data, and indicates that you are in the eighth batch (out of eight).

CHAPTER 6
CONCLUSION

6.1 CONCLUSION

In conclusion, a revolutionary approach to early cancer diagnosis is provided by the creation of a system for pancreatic tumor to detect using Convolutional Neural Networks (CNNs) in medical imaging. Medical procedures could be completely changed by this project's automation of tumor identification, sophisticated feature extraction, and use of data preprocessing techniques, all while taking ethical considerations into account. The system's importance in the field of oncology is highlighted by its potential to improve patient care and outcomes in the fight against pancreatic cancer. It also offers the advantage of faster treatment, improved diagnostic accuracy, and patient privacy protection.

6.2 FUTURE SCOPE

• Expanding the dataset's size and variety Adding more photos from a wider range of patient populations to the dataset will help to enhance the model's performance and generalizability.

• Including more imaging modalities Other imaging modalities, such MRI or PET, may be helpful to include in the study to give a more complete picture of the pancreatic tumor.

REFERENCES

1. Min Li, Xiaohan Nie, Yilidan Reheman, Pan Huang;Shuailei Zhang, Yushuai Yuan, Chen Chen, Ziwei Yan, Cheng Chen, Xiaoyi Lv, Wei Han (2020) Computer aided diagnosis and staging of pancreatic cancer using CT images.

2. Himali Ghorpade, Jayant Jagtap, Shruti Patil, Ketan Kotecha, Ajith Abraham, Natally Horvat, Jayasree Chakraborty (2023) Automatic Segmentation of Pancreas and Pancreatic Tumor: A Review of a Decade of Research.

3. Hong Chen, Gary Y. Hou, Yang Han, Thomas Payen, Carmine F. Palermo, Kenneth P. Olive, Elisa E. Konofagou (2015) Harmonic Motion Imaging for Abdominal Tumor Detection and High-Intensity Focused Ultrasound Ablation Monitoring: An In Vivo Feasibility Study in a Transgenic Mouse Model of Pancreatic Cancer.

4. Sarfaraz Hussein, Pujan Kandel, Candice W. Bolan, Michael B. Wallace, Ulas Bagci (2019) Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches.

5. Koichiro Yasaka , Hiroyuki Akai , Osamu Abe, Shigeru Kiryu et al (2017) Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast enhanced CT.

6. Koichiro Yasaka , Osamu Abe (2017) Pancreatic Ductal Adenocarcinoma Radiology Reporting Template.

7. Atharva Deshpande , Omkar Pedamkar , Nachiket Chaudhary et al (2018) Pancreatic Ductal Adenocarcinoma Radiology Reporting. Statement of the Society of Abdominal Radiology and the American Pancreatic Association.

8. N. Sharma, V. Jain and A. Mishra Et al (2018) An Analysis of Convolutional Neural Networks for Image Classification.

9. Zhang, Y., Huang, J., Liu, C., & Wu, J. Et al (2020) Deep learning for pancreatic tumor detection: A review.

10. R. K. Varghese et al (2021) Review on Deep Learning for Pancreatic Cancer Detection and Diagnosis.

Format - I

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY **(Deemed to be University u/s 3 of UGC Act, 1956)**		
Office of Controller of Examinations		
REPORT FOR PLAGIARISM CHECK ON THE DISSERTATION/PROJECT REPORTS FOR UG/PG PROGRAMMES **(To be attached in the dissertation/ project report)**		
1	Name of the Candidate **(IN BLOCK LETTERS)**	KASANUR JYOTHIRADHITHYA N. MD. ATAHAR HAJEE
2	Address of the Candidate	HOUSING BOARD, ANANTAPUR, ANDHRA PRADESH, 515001 **Mobile No: 7993575800** RAM NAGAR , NELLORE,ANDHRA PRADESH, 515001 **Mobile No: 8074544086**
3	Registration Number	RA2011003010416 RA2011003010892
4	Date of Birth	21 MAY 2003 03 JUNE 2002
5	Department	Computer Science and Engineering (CTECH) Computer Science and Engineering (CTECH)
6	Faculty	School of Computing
7	Title of the Dissertation/Project	COMPUTER AIDED DIAGNOSIS OF PANCREATIC TUMOR BASED ON CT IMAGES
8	Name and address of the Supervisor / Guide	Dr. A. M. J. MUTHUKUMARAN SRM Institute Of Science and Technology, Kattankulathur - 603203 **Mail ID: muthua@srmist.edu.in** **Mobile No:9080583664**
9	Name and address of Co-Supervisor / Co- Guide (if any)	**NILL** **Mail ID:** **Mobile Number:**

11	Software Used	TURNITIN		
12	Date of Verification	09-11-2023		
13	**Plagiarism Details: (to attach the final report from the software)**			
Chapter	**Title of the Chapter**	**Percentage of similarity index (including self citation)**	**Percentage of similarity index (Excluding self-citation)**	**% of plagiarism after excluding Quotes, Bibliography, etc.,**
1	INTRODUCTION	0	1	<1
2	LITERATURE SURVEY	<1	1	2
3	PROJECT DESCRIPTION	<1	<1	<1
4	MODULE DESCRIPTION	<1	<1	<1
5	RESULTS	0	0	0
6	CONCLUSIOND AND FUTURE ENHANCEMENTS	0	1	0
7				
8				
9				
10	TOTAL			7%
Appendices				
I / We declare that the above information have been verified and found true to the best of my / our knowledge.				
Signature of the Candidate		**Name & Signature of the Staff (Who uses the plagiarism check software)**		
Name & Signature of the Supervisor/ Guide		**Name & Signature of the Co-Supervisor/Co-Guide**		
Name & Signature of the HOD				

plag 892 lat

by Muthukumaran AMJ

Submission date: 09-Nov-2023 11:07AM (UTC+0530)
Submission ID: 2221405981
File name: report_RA0416,RA0892_on_9_11.docx (722.72K)
Word count: 5332
Character count: 31007

plag 892 lat

ORIGINALITY REPORT

7%	5%	6%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.science.gov Internet Source	2%
2	www.researchgate.net Internet Source	2%
3	Min Li, Xiaohan Nie, Yilidan Reheman, Pan Huang et al. "Computer-Aided Diagnosis and Staging of Pancreatic Cancer Based on CT Images", IEEE Access, 2020 Publication	1%
4	doaj.org Internet Source	<1%
5	"Proceedings of the 4th International Conference on Data Science, Machine Learning and Applications", Springer Science and Business Media LLC, 2023 Publication	<1%
6	Submitted to King's College Student Paper	<1%
7	ijarsct.co.in Internet Source	<1%

8	Himali Ghorpade, Jayant Jagtap, Shruti Patil, Ketan Kotecha, Ajith Abraham, Natally Horvat, Jayasree Chakraborty. "Automatic Segmentation of Pancreas and Pancreatic Tumor: A Review of a Decade of Research", IEEE Access, 2023 Publication	<1%

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		