Research Article

Advanced and Smart Heart Disease Prediction using Hybrid Machine Learning Techniques

Miss. Srushti N. Raut and Prof. Vikas M. Patil

Department of Computer Engineering Zeal College of Engineering and Research

Abstract

Heart disease is one of the leading cause of death globally because of change in lifestyle of human being. Around 90% of cardiovascular disease can be prevented. Expectation of cardiovascular infection is a basic test in the region of clinical information examination. Health care fields have a vast amount of data, for processing those data certain technologies are used. Machine learning strategies are one of the effective for prediction. Different investigations give just a look into anticipating heart ailment with ML systems. Predict the cause and disease is one of the major challenges now a days. In this paper, we propose a novel strategy that targets finding huge includes by applying AI systems bringing about improving the exactness in the forecast of cardiovascular disease. In Artificial neural network model consist image processing, Image Filtering, Feature Extraction, Segmentation, Edge Detection and Feature Recognition steps. We use Python and Open-CV to make our detector and recognizer. ANN is one of the ML procedures which can be utilized to accomplish productive conclusion results and SVM classifier for getting fusion of results. We are using hybrid machine learning techniques of ANN and SVM for prediction. The proposed framework with ANN and fiveoverlap cross approval are gives 83% grouping exactness with mainly 4 different type of heart disease.

Keywords: ArtificialNeuralNetwork, Heart Disease,Machine Learning

Introduction

Heart disease is the leading Causes of death in the world over the past ten years. It is must that diagnosing system must generate accurate results that proper treatment can be available for the patient. In earlier systems various techniques was used for predicting the reason behind the disease which lacks in accuracy of the output. Various heart diseases like heart attacks, chest pain etc. This thesis is based on the heart disease diagnosis of patients. Heart disease is a prevailing disease nowadays. Now due to increasing expenses of heart disease, there was a need to develop a new system which can predict heart diseases in an easy and cheaper way. The proposed work divides proposed system in two parts such as performance model and prediction model. Performance model is designed to evaluate the overall performance of the application. Prediction Model is used to predict the condition of the patient after evaluation on the basis of various parameters like heart beat rate, blood pressure, cholesterol etc. The accuracy of the system is proved in python.

Businesses (clinics, therapeutic testing focus) is the arrangement of value administrations at reasonable expenses and Quality assistance infers diagnosing patients accurately also, directing medicines that are successful. In the Poor clinical choices can prompt lamentable outcomes which are accordingly unsuitable. Emergency clinics should likewise limit the expense of clinical tests and reports.

They can accomplish these outcomes by utilizing proper PC based data or potentially choice emotionally supportive networks. Diagnosing is a procedure which is finished via robotized frameworks or machines in human framework to discover the idea of the infection by observing the different side effects of the ailment. Diagnosing is a most confused assignment to perform. Diagnosing machines or frameworks hush up accommodating in this procedure in light of the fact that not each specialist must have the information on every single sort of issue of ailment. In this diagnosing physically by specialists can prompt off base outcomes a few times. Accordingly a mechanized diagnosing machine is utilized by them to analyse the issue precisely. The WHO consortium has shared this data that ten a large number of passings happen in this world is a direct result of heart disease. So it was a very risks issue in world. These frameworks regularly produce colossal measures of information which appear as numbers, diagrams and pictures. Tragically, this information is seldom used to help clinical basic leadership.

Objective and scope

1. Main objective behind to develop a system helps the doctors to cross verify their diagnosed results which gives promising solution over existing death rates.
2. By using our proposed work try to invent unique platform and most promising solution for early diagnosis of heart diseases.
3. Existing work analysis accuracy is reduced when the quality of medical data is incomplete. Moreover, different regions exhibit unique characteristics of certain regional diseases, which may weaken the prediction of heart disease wrong.
4. So we are giving more accurate solution by using SVM classifier and artificial neural network (ANN) to detect heart diseases and make predictions using hybrid approach.
5. The scope of this system is to diagnose cardiovascular heart diseases and to make all possible precautions to prevent at early stage itself with affordable rate.

Problem Statement

Heart disease can be managed effectively with a combination of lifestyle changes, medicine and in some cases, surgery. With a right treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The early and accurate prediction result of heart disease is used to prevent it and thus reduce cost for surgical treatment.

Existing work analysis accuracy is reduced when the quality of medical data is incomplete which may weaken the prediction of heart disease wrong. So we are giving more accurate solution by using SVM classifier and artificial neural network (ANN) to detect heart diseases and make predictions using hybrid approach. And also by developing different feature selection method for selecting significant features to increase the performance of heart disease prediction by using diverse mixture of machine learning techniques.

Literature Review

Mamatha Alex P and Shaicy P Shaji [1] state that we are living in a post present day time and there are colossal changes happening to our day by day schedules which make an effect on our wellbeing emphatically and adversely. Because of these progressions different sort of sicknesses are gigantically expanded. Particularly, coronary illness has become increasingly normal these days. The life of individuals is at a hazard. Variety in Blood pressure, sugar, and beat rate and so on can prompt cardiovascular illnesses that incorporate limited or blocked veins. It might cause Heart disappointment, Aneurysm, Peripheral vein infection, Heart assault, Stroke and even abrupt heart failure. Numerous types of heart illness can be distinguished or determined to have distinctive therapeutic tests by thinking about family therapeutic history and different variables. Be that as it may, the forecast of heart maladies without doing any therapeutic tests is very troublesome. The point of this venture is to analyse extraordinary heart ailments and to make every single imaginable safety measure to forestall at beginning period itself with moderate rate. We follow 'Information mining' procedure in which credits are encouraged in to SVM, Random backwoods, KNN, and ANN arrangement Algorithms for the expectation of heart sicknesses. The fundamental readings and studies got from this strategy is utilized to know the probability of identifying heart sicknesses at beginning period and can be totally restored by appropriate determination.

Jyotismita Talukdar, Bhupesh Kumar Dewangan [2] an investigation has been made on the probability and precision of early forecast of a few Coronary illness utilizing Artificial Neural Network. (ANN).The study has been made in both clamour free and uproarious condition. The information gathered for this examination is from five Hospitals. Around 1500 heart patient's information has been gathered and examined. The information is broke down and the results have been contrasted and the Doctor's determination. It is discovered that, in clamor free condition, the precision changes from 74% to 92%.and in loud condition (2dB),the consequences of precision fluctuates from 62% to 82%. In the present examination, four essential traits considered are Blood

Pressure (BP), Fasting Blood Sugar (FBS), Thallic (THAL) and Cholesterol (CHOL).It has been discovered that most noteworthy accuracy(93%), has been acomplished if there should arise an occurrence of PPI(Post Permanent Pacemaker Implementation), around 79% if there should arise an occurrence of CAD(Coronary Artery disease),87% in DCM(Dilated Cardiomyopathy), 89% if there should arise an occurrence of RHD&MS(Rheumatic heart disease with Mitral Stenosis), 75 % if there should arise an occurrence of RBBB +LAFB (Right Bundle Branch Block + Left Anterior Fascicular Block),72% for CHB(Complete Heart Block) and so forth. The most reduced exactness has been acquired if there should arise an occurrence of ICMP (Ischemic Cardiomyopathy),about 38% and AF(Atrial Fibrillation) , around 60 to 62%.

Balasaheb Tarle and Sudarson Jena [3] proposed that definitive point of the proposed strategy is to build up a model for order of medicinal information. In this system we present ANN for Diagnosis of Heart infection. ANN works like the neural course of action of mind. The mind examines from quite a while ago encounters. It can tackle the issues that are not processable or resolvable by current processing frameworks. Highlight determination is most significant procedure for choosing a pertinent property among the colossal informational collection. This decreased arrangement of ascribes is then passed to fake neural system. ANN utilizes back engendering

calculation for better expectation of result, likewise the ANN depends on the information this preferred position can be taken for increasingly precise outcome. ANN gives better precision. The outcome given by ANN is profoundly exact, so it might be philanthropy in the restorative territory for probability of maladies like swine influenza and coronary illness.

Dr.C.Sowmiya and Dr.P.Sumitra [4] introducing Heart disease is the main issue for world. Heart diseases more than individuals passings happen during the principal coronary failure .But not just for cardiovascular failure have a few issues assaulted for bosom malignant growth, lung disease, ventricle. Valve, and so forth… It is basic to have a casing work that can usefully perceive the commonness of coronary illness in thousands of tests in timeously. In this paper the capability of nine (9) characterization strategies was assessed of forecast of coronary illness. To be specific choice tree, innocent Bayesian neural system, SVM.ANN, KNN.My proposed calculation of Apriori calculation and SVM (bolster vector machine) in coronary illness expectation. Utilizing restorative profiles, for example, an age, sex, blood pressure, chest torment type, fasting glucose. It can anticipate like of patients getting coronary illness Based on this, restorative society checks out distinguishing also, forestalling the coronary illness. From the examination it have demonstrated that order based procedures contribute high adequacy and acquire high precision analyse than the past techniques.

Tahira Mahboob et.al [5] proposed Heart maladies may maybe outcome in debility, serious issue, and small nature of life expectancy. Moreover, it could likewise be deadly. Henceforth gathering coronary illness has transform into first misery as of now. This paper focuses on different machine learning rehearses which help determining and seeing countless heart maladies. Diverse AI approaches chatted here are Concealed Markov Models, Support Vector Machine, Highlight Selection, Computational smart classifier, expectation framework, information mining procedures what's more, hereditary calculation. Investigating each approach altogether enabled us to choose most apt one. This at last allows us to propose an Ensemble Model misusing appropriate AI strategies which consummately orders differing heart maladies. The assessment of the proposed method has been led utilizing cutting edge innovation. The proposed strategy has a precision of 94.21%, a ROC (Receiver Operating Characteristics) of 0.981, RMSE (Root Mean Square Error) of .2568, Exactness of 0.953; demonstrating noteworthy improvement when contrasted with the exhibition of K-Nearest Neighbor, Artificial Neural Networks what's more,

Support VectorMachines calculations Investigation/Evaluation of the actualized calculation what's more, the proposed Ensemble Model has been finished consuming the Receiver Operator Characteristics.

Proposed System Approach

This proposed system includes a novel heart detection and prediction mechanism hybrid machine learning Approaches. Which is initially adapts profound highlights and afterward prepares a fake neural system with these scholarly highlights. Exploratory outcomes show the profound educated classifier beats every single other classifier when prepared with all qualities and same preparing tests. It is additionally shown that the presentation improvement is factually huge.

Characterization of heart disease utilizing a low populace, high dimensional informational index is trying because of lacking examples to become familiar with an exact mapping among highlights and class names. In Proposed system consist of hybrid learning approaches with fusion of support vector machine classifiers and Artificial Neural Network. Current working for the most part handles this undertaking through high quality element creation and choice. Deep learning is found to be able to identify the underlying structure of data through the use of ANN and other techniques. This shows that application of machine learning has the potential to significantly detect and classify with almost high accuracy for the low population in India. High dimensional heart disease image data-set is without requiring any hand-crafted, case specific features. High processing speed enhanced ANN classifier model. In our project contains mainly heart disease images type 1, heart disease images type 2, heart disease images type 3, and heart images type 4. In ANN model consist of following stages

- **Image pre-processing :-**

An image is made up of RGB colors. Pre-processing unit consists of noise removal, gray scale conversion, and binary conversion of images by using OTSU's thresholding followed by feature extraction. In future extraction five steps followed in which fingertips searches by eccentricity. Next Elongations of images are measured by considering pixel segmentation as well as rotation of images.

- **Image Enhancement :-**

Filtering is a technique to modify or enhance the image, i.e. to highlight certain features or remove other features. It includes smoothing, sharpening and edge enhancement. Median filter algorithms generate an output pixel by observing the neighborhood of the input pixel in an image. Image filtering algorithms are used to remove different types of noise from the image.
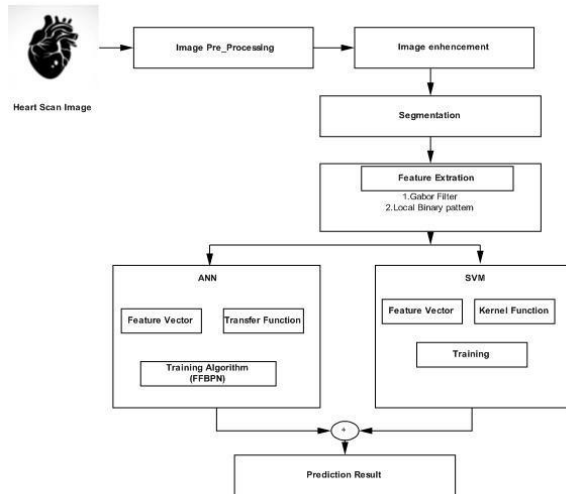
Figure No. 1 System Architecture

- **Segmentation :-**

Image segmentation is the way toward apportioning an advanced picture into various portions (sets of pixels). All pixels in an area share a typical property. Least complex property that pixel can share power. The objective is to disentangle and change the portrayal of the picture into something that is increasingly important and less demanding to break down.

- **Feature Extraction:-**

In feature extraction, algorithmic study used to find the feature vectors of systematic results combine gabor filter and LBP algorithms. In present work "gabor" algorithm which is used to detect image features with greater accuracy. In our system, ANN is used for future recognition in which we having the input unit of training data set of images.

- **Edge Detection :-**

Edge defines the boundaries between regions in an image which helps in object detection. There are many edge detection operators and algorithms available. Edge Detection Operators and Algorithms used in our research like convex hull method.

- **Feature Recognition:-**

Brain-inspired systems used to replicate how humans learn. Consist of input, hidden and output layers that transform the input into something that the output layer can use. Excellent for finding patterns which is complex to human for extract and teach the machine to recognize. ANN gathers their knowledge by detecting the patterns and relationships in data and learns (or is trained) through experience, not from programming.

Mathamatical model
- **System Description:**

S= {I, F, O} **INPUT:**
- F=F1,F2,F3...FN Function to execute result
- **I**=C1,C2,C3... input of systems heart scan images and symptoms reports
- **O**=R1,R2Rn
- I=Result access by User
- C1=Heart disease prediction result **F:**

F1=Image processing applied on heart disease images

F2=feature extraction from images **O:**
R1= model creation from training. R2= model based image testing

SPACE COMPLEXITY:
The space complexity depends on Presentation and visualization of discovered patterns. More the storage of data more is the space complexity.

TIME COMPLEXITY:
We are going to use ANN and SVM for fast and better recognition with higher accuracy. So time complexity is less. So the time complexity of this algorithm is $O(n^n)$.

Success:
1. High accuracy achieved by using heart disease image dataset.
2. User gets result very fast according to their needs.

Failures:
1. Huge data base can lead to more time consumption to get the information.
2. Hardware failure.
3. Software failure.

Mathematical Model in Equation format
Notation
Where, ⮚ M= Set of all entities.
- HDIT1= Heart Disease images type 1
- HDIT2= Heart Disease images type 2
- HDITN= Heart Disease images type N
- HDI=Total Heart Disease images

For calculate total number of heart disease by following equation 1

Total number heart disease images = Total number heart disease images type 1+ Total number heart disease images type 2+..... +Total number heart disease images type N
∑ HDI = ∑ HDIT1+ ∑ HDIT2+...+∑ HDTN......equation 1

Algorithms
Artificial Neural Network (ANN)
ANN calculation utilized in software engineering and other research disciplines, which depends on a huge assortment of basic neural units (fake neurons), freely undifferentiated from the watched conduct of a natural cerebrum's axons. Each neural unit is associated with numerous others, and connections can improve or restrain the initiation condition of bordering neural units. Every individual neural unit figures utilizing plot work. There might be a edge capacity or constraining capacity on every association and on the unit itself, with the end goal that the sign must surpass the point of confinement before spreading to different neurons. These frameworks are self-learning and prepared, instead of unequivocally modified, and exceed expectations in regions where the arrangement or highlight discovery is hard to express in a conventional PC program. Neural systems regularly comprise of numerous layers or a solid shape structure, and the sign way navigates from the main (contribution), to the last (yield) layer of neural units. Back engendering is the utilization of forward incitement to reset loads on

the "front" neural units and this is some of the time done in mix with preparing where the right outcome is known. Dynamic neural systems are the most progressive, in that they powerfully can, based on rules, structure new associations and even new neural units while handicapping others.
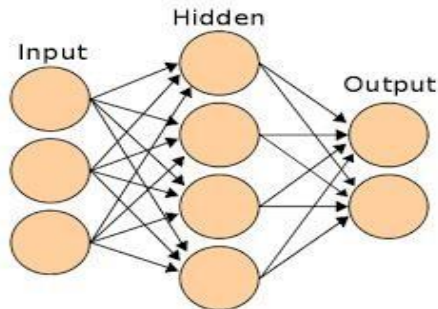

Fig. Layered architecture working of neural network

Training of ANN
The preparation of the ANN is finished after two Passes,

    (I)   A Forward Pass
    (II) a regressive calculation with blunder assurance (III) interfacing weight refreshing in the middle.

It is normal that the preparation must be attempted to quicken the speed of preparing and the rate of meeting of the Mean Square Error (MSE) to the coveted esteem. The consecutive the means are:

1.   Initialization of weight matrix, and
2.   Presenting the Training Samples. In the present study, we have primarily used the Matlab to define a MLP.
3.   To fit with the property of ANN having m-layers of functional units, i.e (m-1) hidden layers.
4.   The Matlab function used is defined as – ffnet=newff(input_data,desired_output {k1, k2 ,......,km−1},{ f1, f2 ,....., fm}, ’’).
5.   The parameters involved with newff( ) are as follows 1. Input Data, [X]: A matrix with input data through its columns in the training set which decides the number of input units for newff( ). 6. Desired Output, [T]: A matrix with the correct answers through its columns in the training set. It decides the number of output units for newff ( ). The commonly used neural network or artificial neural network supervised training algorithm is the Back Propagation Algorithm. The training of a NN/ANN by Back-Propagation algorithm includes the following three stages:

(i)Feed forward algorithm
(ii)Analysis and back propagation of the associated error. (iii)Weight balancing .
In the current analysis of CVD/CHD and their early prediction, we have used the Back-Propagation Algorithm. While training the MLP we have used two training functions:
(i) adapt () (ii) **train ().**
Support Vector Machine (SVM)

It is a discriminative classifier which is formally defined by a separating hyper plane. We can say that the given labeled training data (supervised learning) of heart disease images. SVM is a fast and dependable classification algorithm which performs best with a small and limited amount of heart disease dataset. The idea behind the SVM algorithm is simple to study and predict heart disease, and applying it to its statistical feature vectors of images for classification which doesn't get most of the complex work. And all are the basics of Support Vector Machines. A SVM allows us to classify heart diseases that are actually linearly separable.
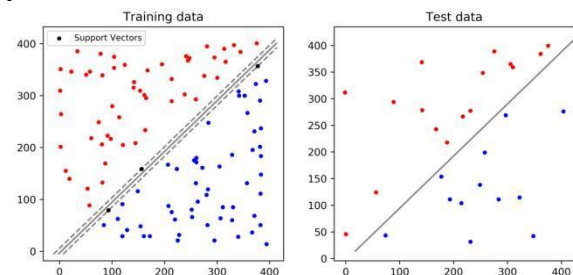

Fig. SVM classifier

Algorithm Steps:
Step1: SVMs maximize the margin around the separating hyper plane.
Assume linear separability for now: in 2 dimensions, can separate by a line in higher dimensions, need hyper planes can find separating hyper plane by linear programming (e.g. perception). Separator can be expressed as :
ax+by=c
Step2: The decision function is fully specified by a subset of training samples, the support vectors.
Step3: the feature vectors generated by using kernel functions
Step4: all features of trained data is used for final classifier to predict results.
Tensor Flow
Machine learning is an unpredictable order. Be that as it may, actualizing AI models is far less overwhelming and troublesome than it used to be, because of AI structures.
Such as Google's Tensor Flow—that facilitate the way toward getting information, preparing models, serving expectations, and refining future outcomes.
Tensor Flow gives the entirety of this to the software engineer by method for the Python language. Python is anything but difficult to learn and work with, and gives advantageous approaches to express how elevated level reflections can be coupled together. Hubs and tensors in Tensor Flow are Python items, and Tensor Flow applications are themselves Python applications.
Open-CV
    •   Monty Python's Flying Circus had a "cat detector van" so, in this tutorial, we use Python and Open-CV to make our very own cat detector and recognizer. We also cover examples of human face detection and recognition. More

generally, we cover a methodology that applies to training a detector (based on Haar cascades) for any class of object and a recognizer (based on LBPH, Fisher-faces, or Eigen-faces) for any unique objects. We build a small GUI app that enables an LBPH-based recognizer to learn new objects interactively in real time. Although this tutorial uses Python, the project could be ported to Android and IOS using Open-CV's Java and C++ bindings.

• Attendees will pick up involvement with utilizing OpenCV to distinguish and perceive visual subjects, particularly human and creature faces. GUI improvement will likewise be stressed. Participants will be guided toward extra data in books and on the web. There is no proper assessment of participants' work yet participants are welcome to exhibit their work and talk about the outcomes they have accomplished during the session by utilizing various finders and recognizers and various parameters.

Python

• Python interface is being actively developed right now. There are many algorithms and many functions that compose or support those algorithms. Open CV is written natively in C++ and has a template interface that works seamlessly with STL containers.

Image Processing

• Read and Write heart disease images. Image processing is basic step to preprocess data in desire format and scale. Final result images get transformed from one form to other is done by using Open-CV library. The image gets processed by using predefined image processing algorithms. The processed images get trained by using train algorithms ANN and SVM.

Comparative Results

In our experimental setup, in table no.8.1, shows total number of 1500 images of heart disease. In our project contains mainly heart disease images type 1, heart disease images type 2, heart disease images type 3, and heart images type 4. In this project consist of 375 numbers of images of heart disease type 1, 425 numbers of images of heart disease type 2, 225 numbers of images of heart disease type 3 and 475 numbers of images of heart disease type 4.

Sr. No	Types	Number of images
1	Heart disease images type 1	375
2	Heart disease images type 2	425
3	Heart disease images type 3	225
4	Heart disease images type 4	475

Table. Heart disease image testing

Results

From above data, as shown in graph 9.1, the total numbers of images of heart disease type 1 were 375, total numbers of images of heart disease type 2 were 425, total numbers of images of heart disease type 3 were 225 and total numbers of images of heart disease type 3 were 475.
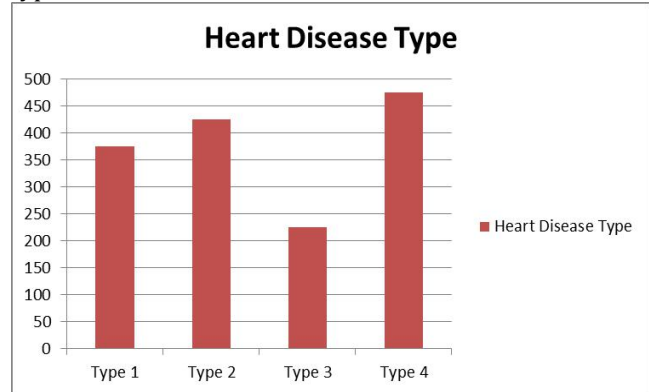
Graph 9.1: Total number of heart disease type

Conclusion

In this technique, we have anticipated heart disease detection by using fusion of machine learning techniques ANN and SVM for accurate prediction. We use Python and Open-CV to make our detector easy and efficient recognizer. Excellent for finding features which is complex to extract and construct the model to recognize. Our proposed technique most accurate and builds exactness over existing work by using deep neural approach. We trust this might be contributed towards reliable and noble heart disease. By this framework adequately if extra examination will empowered in future.

Future Work

Future work will be based on real time heart disease image dataset with considering high accuracy over proposed work.

Acknowledgment

required infrastructure and support. Finally, we would like to extend a heartfelt gratitude to friends and family members.

References

[1] Mamatha Alex P and Shaicy P Shaji "Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique" International Conference on Communication and Signal Processing, April 4-6, 2019, India

[2] Jyotismita Talukdar, Bhupesh Kumar Dewangan, "Analysis of cardiovascular diseases using artificial neural network" 5th IEEE International Conference on Parallel, Distributed and Grid Computing(PDGC-2018), 20-22 Dec, 2018, Solan, India

[3] Balasaheb Tarle and Sudarson Jena "An Artificial Neural Network Based Pattern Classification Algorithm for Diagnosis of Heart Disease" 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA) 2017

[4] C.Sowmiya and Dr.P.Sumitra, "Analytical Study of Heart Disease Diagnosis Using Classification Techniques." 2017 IEEE international conference on intelligent techniques in control, optimization and signal processing

[5] Abbas Khosravi, Amin Khatami, "Lung cancer classification using deep learned features on low population dataset." Canadian Conference on Electrical and Computer Engineering (CCECE) 978-1-5090-5538-8/17 2017 IEEE.